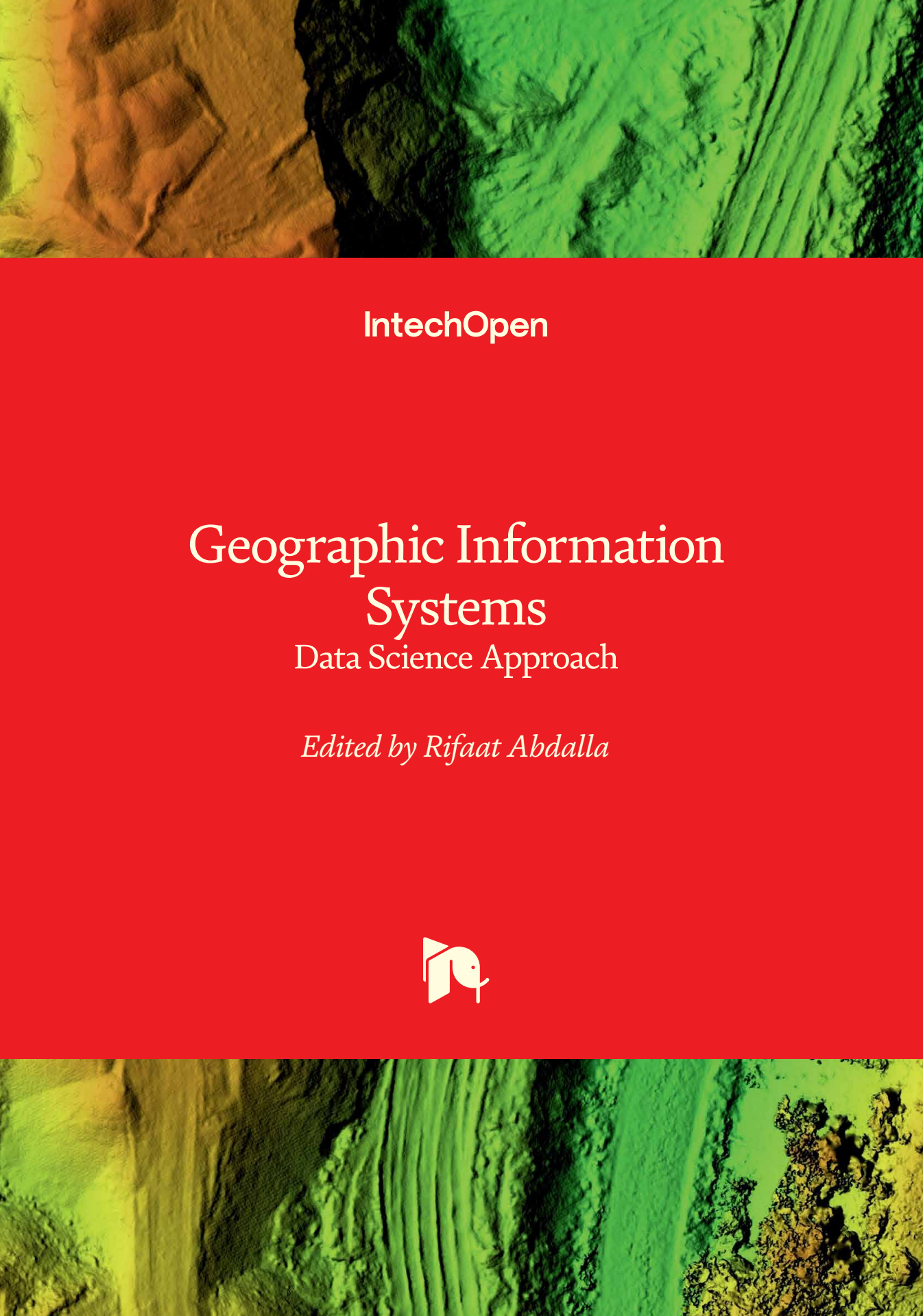




IntechOpen

Geographic Information Systems

Data Science Approach

Edited by Rifaat Abdalla



Geographic Information Systems - Data Science Approach

Edited by Rifaat Abdalla

Published in London, United Kingdom

Geographic Information Systems – Data Science Approach

<http://dx.doi.org/10.5772/intechopen.111053>

Edited by Rifaat Abdalla

Contributors

Emad H. E. Yasin, Czimmer Kornel, Reda Alkot Mohamed, Sayed Ahmed Sayed, Rifaat Abdalla, Imen Hamdi Nasr, Mohamed B. O. Osman, Georgios Zagalikis, Anwarelsadat Eltayeb Elmahal, Mohammed Mahmoud Ibrahim Musa, David Barilla, Giuseppe Caristi, Roberto Guarneri, Dario Lo Bosco, Melgris José Becerra, Mariano Araujo Bernardino da Rocha, Hafiz Usman Ahmed Khan

© The Editor(s) and the Author(s) 2024

The rights of the editor(s) and the author(s) have been asserted in accordance with the Copyright, Designs and Patents Act 1988. All rights to the book as a whole are reserved by INTECHOPEN LIMITED. The book as a whole (compilation) cannot be reproduced, distributed or used for commercial or non-commercial purposes without INTECHOPEN LIMITED's written permission. Enquiries concerning the use of the book should be directed to INTECHOPEN LIMITED rights and permissions department (permissions@intechopen.com).

Violations are liable to prosecution under the governing Copyright Law.



Individual chapters of this publication are distributed under the terms of the Creative Commons Attribution 3.0 Unported License which permits commercial use, distribution and reproduction of the individual chapters, provided the original author(s) and source publication are appropriately acknowledged. If so indicated, certain images may not be included under the Creative Commons license. In such cases users will need to obtain permission from the license holder to reproduce the material. More details and guidelines concerning content reuse and adaptation can be found at <http://www.intechopen.com/copyright-policy.html>.

Notice

Statements and opinions expressed in the chapters are those of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

First published in London, United Kingdom, 2024 by IntechOpen

IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, 5 Princes Gate Court, London, SW7 2QJ, United Kingdom

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

Additional hard and PDF copies can be obtained from orders@intechopen.com

Geographic Information Systems – Data Science Approach

Edited by Rifaat Abdalla

p. cm.

Print ISBN 978-1-83769-868-4

Online ISBN 978-1-83769-867-7

eBook (PDF) ISBN 978-1-83769-869-1

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900+

Open access books available

185,000+

International authors and editors

200M+

Downloads

156

Countries delivered to

Our authors are among the
Top 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Meet the editor



Dr. Rifaat Abdalla, PGeo, GISP, PMP, PMI-RMP, is an esteemed professor celebrated for his exceptional contributions to Geographic Information Systems (GIS) and Remote Sensing. His illustrious career spans academia, government, and industry, earning him accolades and awards for his remarkable work. Notably, he received the esteemed 2008 ASPRS Best Paper award, a prestigious recognition in geomatics, showcasing his outstanding contributions to the field. In 2022, Dr. Abdalla was honored with the Outstanding Leadership Award at the Education2 Conference in Dubai, highlighting his exceptional leadership skills and mentorship in GIS and Remote Sensing. As a Professional Geoscientist (PGeo), Geographic Information Systems Professional (GISP), Project Management Professional (PMP), and Risk Management Professional (PMI-RMP), Dr. Abdalla exemplifies excellence in his field. His expertise and credentials testify to his impactful role in advancing GIS and remote sensing through pioneering research and teaching. Dr. Abdalla's work inspires future generations, cementing his influential legacy in the geospatial community.

Contents

Preface	XI
Acknowledgment	XIII
Section 1	
Data-Driven Geospatial Innovations	1
Chapter 1	3
Evaluating Satellite Image Classification: Exploring Methods and Techniques <i>by Emad H.E. Yasin and Czimmer Kornel</i>	
Chapter 2	37
Programming an Interactive Map to Enhance Clients' Participation in Site Selection of E-Marketing in Cairo and Giza Governorates: Egypt <i>by Reda Alkot Mohamed and Sayed Ahmed Sayed</i>	
Chapter 3	55
Visualization and Machine Learning Prediction of Spatiotemporal Spread of COVID-19 in India <i>by Rifaat Abdalla and Imen Hamdi Nasr</i>	
Chapter 4	75
The Synergy of Data Science, GIS Spatial Analysis and Knowledge Management as a Path to Sustainability Insights <i>by Mohamed B.O. Osman</i>	
Chapter 5	93
Remote Sensing and GIS Applications in Wildfires <i>by Georgios Zagalikis</i>	
Section 2	
Data Science in Geospatial Analysis and Applications	137
Chapter 6	139
Spatial Cloud Computing Using Google Earth Engine and R Packages <i>by Anwarelsadat Eltayeb Elmahal and Mohammed Mahmoud Ibrahim Musa</i>	

Chapter 7	161
A Probabilistic Approach for Road Network Analysis	
<i>by David Barilla, Giuseppe Caristi, Roberto Guarneri and Dario Lo Bosco</i>	
Chapter 8	185
Applications of Geotechnologies in the Field of Public Health	
<i>by Melgris José Becerra and Mariano Araujo Bernardino da Rocha</i>	
Chapter 9	213
Land Use and Land Cover Simulation	
<i>by Hafiz Usman Ahmed Khan</i>	

Preface

Data science and Geographic Information Systems (GIS) share a symbiotic relationship, with data science providing the analytical techniques and methodologies to extract insights from spatial data managed and processed by GIS. GIS generates and organizes geospatial information, while data science employs statistical analysis, machine learning, and visualization tools to derive patterns, trends, and valuable insights from this geographically referenced data. The fusion of these disciplines enables a deeper understanding of spatial relationships, aiding in decision-making processes across various domains like urban planning, environmental management, disaster response, and beyond, thus emphasizing the inherent synergy between data science and GIS in unlocking actionable intelligence from spatial data.

An open-access, edited volume on GIS utilizing a data science approach caters to a wide audience for several reasons. Firstly, it promotes accessibility, allowing researchers, practitioners, and students worldwide to access valuable content without financial constraints. By providing open access, it encourages collaboration and knowledge sharing among experts in GIS and data science, fostering innovation and advancements in these fields. Additionally, it serves as a comprehensive resource, offering diverse perspectives, methodologies, and case studies that showcase the integration of GIS and data science, further stimulating interdisciplinary research and applications. Ultimately, an open-access book facilitates the dissemination of crucial information, fostering a more inclusive and informed community dedicated to leveraging GIS with data science methodologies.

Geographic Information Systems – Data Science Approach encompasses a comprehensive range of topics at the intersection of GIS and data science, providing a holistic view of their applications and integration. It begins by delving into the convergence of machine learning and GIS, showcasing how advanced algorithms enhance geographical analysis. The synergy between data science, GIS spatial analysis, and knowledge management is highlighted, emphasizing their collaborative role in achieving sustainability insights. The book further explores spatial cloud computing using platforms like Google Earth Engine and R Packages, enabling scalable and efficient geospatial data processing. It covers satellite image classification methodologies, focusing on evaluating diverse techniques for accurate remote sensing applications. Additionally, it discusses land use and land cover simulation, crucial for understanding dynamic environmental changes. Integration of GIS and remote sensing for gully erosion assessment showcases the practical utility of these technologies in environmental studies. The probabilistic approach for network analysis emphasizes the role of GIS in network modeling and optimization. Moreover, the book explores the practical application of geotechnologies in public health and presents a case study on programming interactive maps for enhancing client participation in e-marketing site selection in Cairo, Giza Governate, Egypt. Finally, it concludes by illustrating the significance of visualization and machine learning in predicting the spatiotemporal spread of diseases like Covid-19, offering insights into predictive modeling within a geographical

context. Overall, this book presents a rich tapestry of methodologies, applications, and case studies demonstrating the fusion of GIS and data science across various domains.

In closing, it is my sincere hope that the diverse array of topics and insights encapsulated within this book, *Geographic Information Systems – Data Science Approach*, will serve as an invaluable resource to its readers. May this collection of chapters, spanning from the convergence of machine learning and GIS to the practical applications in public health, environmental studies, and beyond, inspire innovation and foster deeper understanding within the realms of GIS and data science. It is my wish that the amalgamation of methodologies, case studies, and interdisciplinary approaches presented here will not only prove to be informative but also catalyze new ideas and advancements in this evolving field. May each reader find practical utility and inspiration within these pages, contributing meaningfully to the collective domain knowledge and application of GIS and data science for the betterment of our global community.

Rifaat Abdalla

Department of Earth Sciences,
College of Science,
Sultan Qaboos University,
Al Khoudh, Oman

Acknowledgment

“For my enduring colleagues who imparted the essence of a peaceful existence, and to my exceptional companions who illuminated life’s profound values. To my mentors, whose unwavering guidance ignited the spark within, urging me to pursue excellence. This book is a tribute to your invaluable teachings and inspiration.”

Section 1

Data-Driven Geospatial Innovations

Evaluating Satellite Image Classification: Exploring Methods and Techniques

Emad H.E. Yasin and Czimber Kornel

Abstract

Satellite image classification serves a critical function across various applications, from land cover mapping and urban planning to environmental monitoring and disaster management. In recent years, significant advancements in machine learning and computer vision, coupled with increased accessibility to satellite imagery, have driven considerable progress in this field. Classification techniques for satellite imagery can be primarily divided into three key approaches: automatic, manual, and hybrid. Each approach offers unique advantages but also comes with its own set of limitations. While most methodologies gravitate toward automatic techniques, choosing an appropriate method should be a carefully considered decision based on specific needs. This paper provides an exhaustive review of cutting-edge classification algorithms, including Artificial Neural Networks (ANNs), Classification Trees (CTs), and Support Vector Machines (SVMs). It also offers a comparative analysis between these modern methods and traditional techniques, focusing on their respective performance metrics when applied to satellite data. This study examines key factors affecting remote sensing data classification, including classifier parameter adjustments and combining multiple classifiers. It reviews existing literature to enhance feature selection and classifier optimization for better accuracy. However, it also points out the continuous need for research in image processing to improve classification accuracy.

Keywords: satellite image classification, remote sensing, classification techniques, machine learning, summary of reviews

1. Introduction

Satellite image classification is a crucial element in the fields of remote sensing and Earth observation. It has a wide array of applications, including but not limited to land cover mapping, environmental monitoring, disaster management, urban planning, and agricultural assessment [1]. Essentially, the classification process categorizes pixels or regions within satellite images into predefined classes or land cover types. This enables the extraction of vital information about the Earth's ever-changing features [2], serving as a foundation for data-driven decisions across various sectors—

from climate change mitigation and resource management to humanitarian efforts [1]. Recent advancements in technology have dramatically transformed the landscape of satellite image classification. The introduction of high-resolution satellites and the increasing availability of multispectral and hyperspectral data have set the stage for a new era in this field [3]. These advances allow for the incorporation of sophisticated machine learning and deep learning algorithms, which substantially enhance feature extraction and enable more context-sensitive classifications [3]. In this evolving context, the critical evaluation of classification methodologies has gained heightened importance. Robust and reliable evaluation frameworks are indispensable for assessing the performance of various classification models, ensuring that the insights garnered are both accurate and actionable [3]. This paper aims to offer a thorough analysis of the evaluation methodologies and techniques currently prevalent in the field. It also provides an overview of contemporary trends, challenges, and innovations in satellite image classification [4]. By examining existing studies, we aim to present a comprehensive perspective on cutting-edge methods, unresolved challenges, and emerging directions in evaluation techniques within this domain. The primary goal of this review is to assist analysts, especially those new to the dynamic world of remote sensing, in making well-informed choices when selecting appropriate classification methods. This will, in turn, lead to more accurate satellite image classifications and the generation of precise Land Use and Land Cover (LULC) maps [3]. In particular, this review focuses on groundbreaking advances in classification algorithms, especially Artificial Neural Networks (ANNs), Classification Trees (CTs), and Support Vector Machines (SVMs). These state-of-the-art techniques have shown considerable promise in the realm of satellite image classification, opening new pathways for enhancing both accuracy and robustness.

1.1 Setting the context for satellite image classification

Satellite image classification sits at the intersection of remote-sensing technology, artificial intelligence, and Geographic Information Systems (GIS). With Earth-monitoring satellites generating vast quantities of data daily, the need for automated techniques to extract valuable insights from this imagery has never been more pressing. These images encompass a range of spectral bands—including optical, infrared, and radar—all of which provide unique and comprehensive data on Earth's terrestrial and aquatic features. The primary objective of satellite image classification is to transform raw satellite data into actionable intelligence. This involves the identification and categorization of various objects or land cover types, such as forests, urban areas, water bodies, and agricultural fields. This categorization serves as a vital foundation for numerous applications in diverse fields. Whether the focus is on urban planning, precision agriculture, emergency response, or environmental monitoring, the insights derived from accurate satellite image classification are invaluable.

1.2 Satellite image classification techniques

Satellite images can typically be classified into three main categories based on spatial resolution: low-, medium-, and high-resolution. These categories are determined by the pixel size or granularity of the images, which directly influences both the level of detail captured and the suitability of the images for various applications (see **Figure 1**). Low-resolution images feature coarser pixels and, as a result, cover



Figure 1.
High, medium and low spatial resolution satellite images.

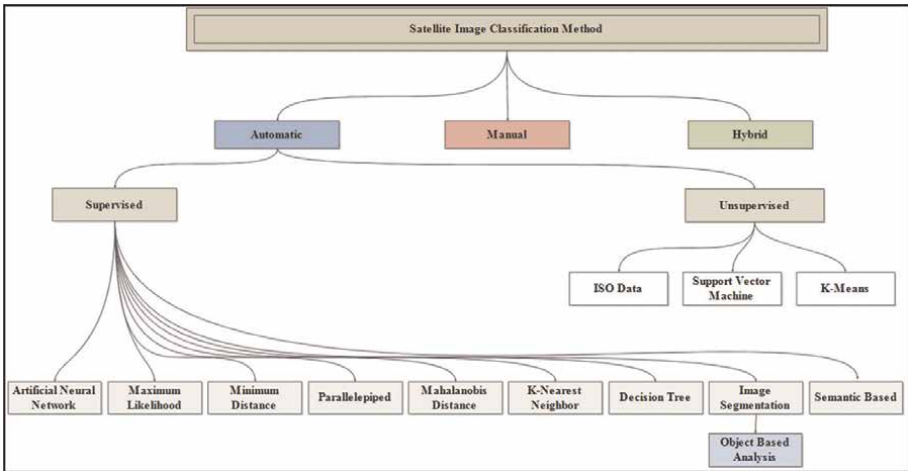


Figure 2.
Hierarchy of satellite image classification techniques [5].

extensive geographic areas but provide limited detail. Medium-resolution images offer a middle ground, capturing a moderate level of detail while still encompassing relatively large areas. In contrast, high-resolution images are characterized by fine pixels that capture intricate details, making them ideal for specialized and precise analyses. Understanding these distinctions is crucial for selecting the most appropriate type of satellite imagery to meet specific remote-sensing goals and geospatial analysis requirements.

The classification of satellite images is organized into a diverse array of methods and techniques, offering a structured approach to data analysis (refer to **Figure 2**). Typically, these methods fall under three major categories, each with its own distinctive features and use cases: A. Manual classification, B. Automatic classification, and C. Hybrid classification.

1.2.1 Manual classification

Manual classification techniques are esteemed for their reliability, efficiency, and effectiveness. They rely heavily on the interpretative skills of analysts who leverage their extensive knowledge of local conditions and the unique characteristics of the study area. However, it is crucial to recognize that manual classification can be resource-intensive, often requiring the expertise of field specialists for optimal accuracy. In this approach, the quality and reliability of the classified data are intrinsically linked to the analyst's skill set and familiarity with the specific area of study.

1.2.2 Automatic classification

While manual classification is closely tied to the expertise of the analyst, automatic classification methods offer an alternative that mitigates this variable. These techniques categorize pixels based on quantifiable metrics like similarity and dissimilarity. Depending on the spatial resolution of the images, automatic classification can be broken down into three distinct categories: (a) pixel-based classification, which focuses on individual pixels; (b) sub-pixel-based classification, which examines fractions of a pixel; and (c) object-based classification, which considers groups of pixels as single entities.

1.2.2.1 Pixel-based classification

Pixel-based techniques form the bedrock of remote-sensing image analysis. These methods operate under the assumption that each pixel in an image corresponds to a uniform land use or land cover type [6]. Within this framework, remote-sensing data are treated as an array of unique pixels, each defined by specific spectral features and associated variables. Generally speaking, pixel-based classification can be divided into two principal groups:

1.2.2.1.1 Unsupervised classification

Unsupervised classification techniques segment an image into distinct classes based on the natural clustering of pixel values, without requiring prior knowledge or training data. Notable algorithms in this category include K-Means and its variant, ISODATA (Iterative Self-Organizing Data Analysis Technique). More recent developments have adapted algorithms like Support Vector Machines (SVMs) and hierarchical clustering methods for unsupervised purposes [7]. However, these methods come with their own set of challenges, such as computational intensity, particularly when handling large datasets. Another limitation is the often-limited accuracy in generating meaningful and specific class distinctions, which may require additional post-processing or user intervention.

1.2.2.1.2 Supervised classification

In supervised classification, as illustrated in **Figure 3**, satellite images are classified using predefined input or training data. An analyst carefully selects representative areas with known class types, commonly referred to as training samples or signatures. The spectral characteristics of individual pixels within the image are then compared against these training samples. Based on predefined decision rules, the analyst categorizes each pixel into the most appropriate class [8]. Over the years, a diverse array

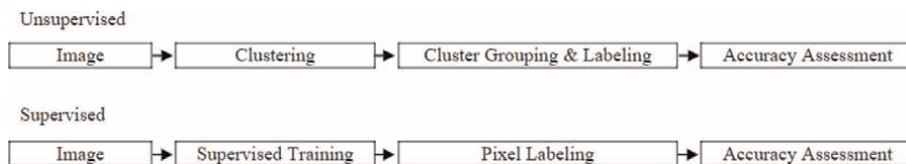


Figure 3.
 The major steps of supervised and unsupervised image classification.

of supervised classification techniques have emerged, enriching the framework for satellite image classification, as further detailed in **Figure 3**.

1.2.2.2 Sub-pixel-based classification

Traditional pixel-based methods commonly operate on the premise that each pixel in an image exclusively represents one type of land use or land cover. This assumption becomes less sustainable when working with medium or coarse spatial resolution satellite images due to the inherent heterogeneity of natural landscapes [9]. Sub-pixel-based classification techniques present a more nuanced solution by precisely quantifying the distribution of various land use or land cover types within each pixel [10]. Leading techniques in this category include fuzzy classification, neural networks [11], regression modeling [12], regression tree analysis [13], and spectral mixture analysis [14]. Engineered to address the challenges of heterogeneous pixel compositions, these methods offer a refined way to capture the diversity of land use and land cover types within a single pixel.

1.2.2.3 Object-based classification

Distinct from pixel-based and sub-pixel-based methods, object-based classification offers a novel paradigm for analyzing remote-sensing imagery [1]. Rather than viewing individual pixels as the elemental units for classification, this approach concentrates on geographical objects—formed through image segmentation [15]—as the foundational units. This methodology is especially advantageous for very high-resolution (VHR) remote-sensing images. Object-based classification integrates spectral, spatial, and contextual information through segmentation techniques, producing more comprehensive image objects for analysis. The central assumption driving this approach is that geographical objects consist of multiple components within an image, thereby enabling a more intricate and accurate representation of complex scenes. Empirical evidence consistently indicates that object-based methods yield substantially higher accuracy rates compared to pixel-based alternatives (**Figure 4**) [17].

1.2.3 Hybrid classification

The object-based classification approach blends aspects of both manual and automated methods to forge a hybrid system. Analysts can guide the classification process by offering initial input, refining interim results, or injecting domain-specific expertise. At the same time, machine-learning algorithms can be integrated to automate segments of the classification process, enhancing efficiency. This hybrid methodology proves especially beneficial in contexts where the nuanced interpretation provided by human expertise is invaluable, but the speed and scalability offered by automation are

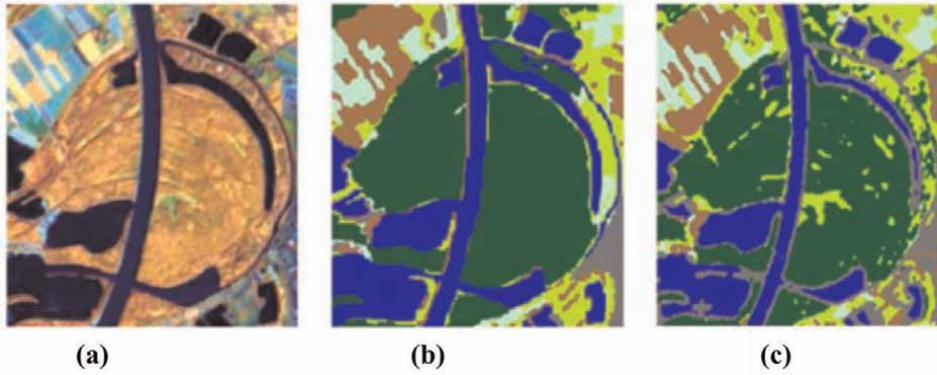


Figure 4. (a) Satellite image, (b) object-based classification of a satellite image and (c) pixel-based classification of a satellite image [16].

also key. Consequently, it is frequently utilized in situations where a mix of manual and automated tactics delivers the best results.

2. Satellite image classification methods

Satellite image classification methods serve as indispensable tools for extracting valuable insights from satellite imagery. They enable the categorization of various features, including land cover and land use, within these images. This section outlines the latest advancements in satellite image classification techniques:

2.1 K-means

K-means clustering, a well-known iterative technique originating from data mining, falls under the category of unsupervised classification known as cluster analysis [18]. In this approach, a set of n observations is divided into k clusters. Each observation is allocated to the cluster to which its mean is the closest. The process starts with the assignment of a random initial cluster vector. Following this, each pixel is classified into the cluster closest to it in terms of distance. The cluster means are then recalculated based on all the pixels contained within each cluster. These latter two steps are repeated until the mean values of the clusters remain stable, effectively minimizing within-cluster variability and leading to cluster convergence. The Sum of Squares Distances function, expressed in Eq. (1), quantifies the relationship between each pixel in the dataset and the centroid (mean) of its designated cluster. This function calculates the squared Euclidean distance between each pixel and the cluster center to which it belongs. This measure is instrumental in determining how closely or loosely each pixel aligns with its cluster centroid, thus shedding light on the level of homogeneity within the clusters as well as the overall efficiency of the clustering process. The ultimate aim of clustering algorithms like K-means is to minimize this Sum of Squares Distances function. By doing so, the algorithm ensures that pixels within the same cluster are as similar as possible while maximizing dissimilarity between distinct clusters. Achieving this goal results in the formation of compact and well-separated clusters, thereby improving the quality of the unsupervised classification outcomes.

$$SS_{\text{dist}} = \sum_{x \in c} [x - c(x)]^2 \quad (1)$$

where $c(x)$ is the mean of the cluster pixel to which x is assigned.

Minimizing the Mean Squared Error (MSE) is closely linked to minimizing the Sum of Squares Distances (SSdist), as denoted in Eq. (2). The MSE serves as a significant metric for assessing the variability within clusters. It calculates the average of the squared differences between the position of each pixel and the centroid of the cluster to which it belongs. In the context of clustering, a lower MSE value suggests that the pixels within each cluster are tightly grouped around their respective centroids, indicating lower variability and higher internal cohesion. Thus, by aiming to minimize the MSE, one also effectively contributes to the reduction of the SSdist. Achieving this dual minimization helps in creating more compact and well-defined clusters, fulfilling a fundamental goal in both cluster analysis and unsupervised classification methods. This approach enhances the quality of the unsupervised classification output by ensuring that pixels within the same cluster are as similar as possible, thereby also maximizing the dissimilarity between different clusters.

$$MSE = \frac{\sum_{x \in c} [x - C(x)]^2}{(N - c)^b} = \frac{SS_{\text{dist}}}{(N - c)^b} \quad (2)$$

2.2 Iterative self organizing data (ISODATA)

As outlined by [19], the ISODATA algorithm is equipped with a dynamic adjustment mechanism that continually refines the cluster set through a series of iterative calculations. This fine-tuning is achieved primarily through two key processes: cluster assimilation and cluster splitting, visualized in **Figure 5**. Cluster assimilation is activated under specific conditions: either when the number of pixels within a cluster falls below a predetermined minimum threshold or when the distance between the centroids of two clusters is less than a specified value. In such cases, the algorithm opts to merge or “assimilate” the clusters into a unified group. On the other hand, cluster splitting is initiated when a cluster’s internal standard deviation exceeds a set threshold and when the number of pixels within that cluster is at least double the predetermined minimum. Under these conditions, the algorithm divides or “splits” the existing cluster into two or more separate clusters. These dynamic operations

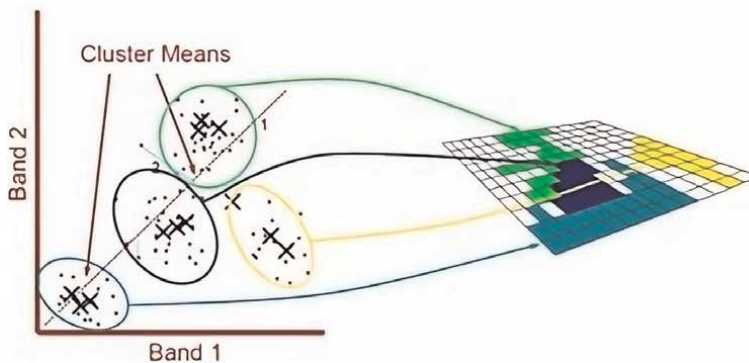


Figure 5.
 ISODATA classification pixels using cluster means [16].

enable the ISODATA algorithm to adjust responsively to the intrinsic variability and complexity present within the dataset. Such adaptability ensures that the resulting clusters are a reliable and precise reflection of the inherent patterns and structures in the data, thereby enhancing the quality of the classification.

The ISODATA algorithm stands out for its adaptability in managing datasets that feature a variable number of clusters. Unlike some other clustering algorithms that require a pre-specified number of clusters, ISODATA dynamically adjusts this number during the clustering process. It does so based on the specific traits of the data and the guidelines set for cluster assimilation and splitting, as previously discussed [20]. This feature adds an extra layer of flexibility, making the algorithm well-suited for handling complex datasets where the optimal number of clusters is not easily determinable beforehand.

2.3 Support Vector Machine (SVM)

As outlined by [21], the Support Vector Machine (SVM) is grounded in statistical learning theory and is primarily used for classification tasks. It works by partitioning the data into separate classes through a hyperplane designed to maximize the gap, or margin, between these classes. In this context, data points that lie close to the hyperplane are identified as “support vectors,” as illustrated in **Figure 6**. SVM aims to optimize this margin to reduce the probability of misclassification errors [18]. In achieving this goal, SVM establishes a distinct and unambiguous boundary separating various classes in the dataset.

A Binary classification of N training samples and for each instance, consisting of a tuple (x_i, y_i) ($i = 1, 2, \dots, N$) where, $x_i = (x_{i1}, x_{i2}, \dots, x_{id})^T$ deals with the attribute set for the i th sample $\in \{-1, +1\}$, and let y_i denote the class labels.

The decision boundary of a linear classifier, in the context of Support Vector Machines (SVM), can be mathematically expressed as:

$$\vec{W} \cdot \vec{X} + b = 0 \quad (3)$$

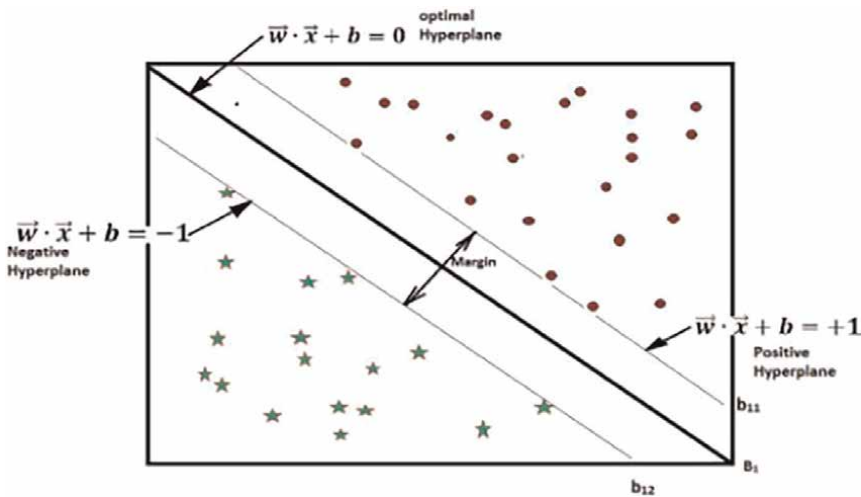


Figure 6.
Margin of decision boundary in binary SVM classifier [16].

In a binary classification scenario, as is commonly used with Support Vector Machines (SVM), distinct types of data points are given specific class labels for training purposes. For example, if circles are assigned a class label of $+1$ and stars are given a class label of -1 , this labeling forms the basis for the classification algorithm. In this context, the class label y for a test sample t can be determined using the following decision rule:

$$y = \begin{cases} +1 & \text{if } \vec{W} \cdot \vec{t} + b > 0; \\ -1 & \text{if } \vec{W} \cdot \vec{t} + b < 0. \end{cases} \quad (4)$$

The distance between these two hyperplanes gives the marginal decision boundary (d) (see Eq. (5))

$$d = \frac{2}{\|\vec{w}\|} \quad (5)$$

2.4 Minimum Distance Classification

In the Minimum Distance to Means (MDTM) technique, which is a form of supervised classification, a single mean vector that captures its essential spectral characteristics represents each class. For any given pixel in the image, its measurement vector is compared to the mean vectors of each class to calculate spectral distances. Based on these distances, a decision rule, as depicted in **Figure 7**, is then applied to classify the pixel into one of the predefined classes. The strength of the MDTM classifier lies in its simplicity and computational efficiency, especially when a singular mean vector can adequately represent each class. By directly calculating the spectral distance between a candidate pixel and the mean vectors of each class, the MDTM method ensures that the pixel is classified into the category to which it is most spectrally similar. This approach is particularly useful in scenarios where computational resources are limited but a reasonable level of classification accuracy is still required.

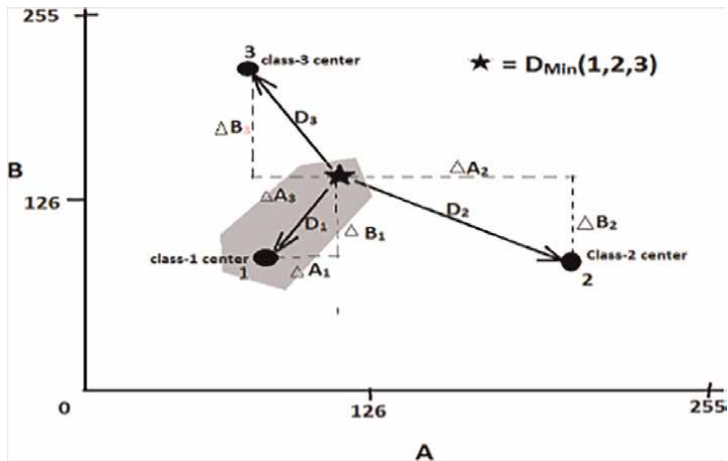


Figure 7.
Calculation of minimum distance between centers of three classes and candidate pixel concerning bands A & B [16].

As specified by Eq. (6), the classifier uses Euclidean distance as the means to calculate the gap between a candidate pixel's measurement vector and the mean vector of each predefined class. After performing this calculation, the candidate pixel is allocated to the class with the smallest Euclidean distance. This approach leverages the Euclidean distance as a numerical metric for determining the class most spectrally akin to each candidate pixel. By aiming to minimize this distance, the algorithm accurately places the candidate pixel into its most probable class, thereby improving the overall precision of the classification.

$$D_{ab} = \sum_{i=0}^n \left((a_i - b_i)^2 \right)^{1/2} \quad (6)$$

where D_{ab} = distance between class a and pixel b , a_i = mean spectral value of class a in band i , b_i = spectral value of pixel b in band i , and n = number of spectral bands.

2.5 Mahalanobis Distance classifier

As highlighted by [13], the Mahalanobis Distance classifier shares similarities with the Minimum Distance to Means technique, but it adds a key element: the use of a covariance matrix in the classification of satellite images. This matrix captures the statistical interrelationships among various spectral bands, providing a more nuanced and multidimensional metric for measuring the distance between data points than the Euclidean distance used in the Minimum Distance approach. Importantly, the integration of the covariance matrix renders the Mahalanobis Distance classifier particularly effective in situations where the intra-class data distribution exhibits variable levels of variance across different dimensions. As a result, the classifier is more adept at handling the complex spectral interdependencies present in the data, thereby improving the overall accuracy and dependability of the classification.

$$\text{Mahalanobis Distance } (Dx) = \frac{(\mathbf{x} - \mu_i)^T}{\sum_i} (\mathbf{X} - \mu_i) \quad (7)$$

Where, $\sum$ = pixel covariance matrix for class i , ($i = 1, 2, \dots, n$), μ = average vector of class i

2.6 Parallel Piped classification

As explained by [15], the Parallel Piped classifier divides each axis of the multi-spectral attribute space into specific regions, often called classes. The limits of these regions are set by the lowest and highest pixel values along each corresponding axis. The success of this classifier in terms of precision and effectiveness largely depends on a thorough analysis of the range for each class, guided by relevant population statistics (as shown in **Figure 8**). By carefully assessing these ranges, the Parallel Piped classifier strives to accurately allocate pixels to their appropriate classes, thus improving the overall quality and trustworthiness of the image classification process.

In a 2D feature space, the Parallel Piped classifier creates a set of rectangular boxes, each signifying a unique class. Pixels that fall within a particular box are then classified as members of the associated class. This technique is notable for its computational simplicity, as it speeds up the classification process by concentrating on predefined

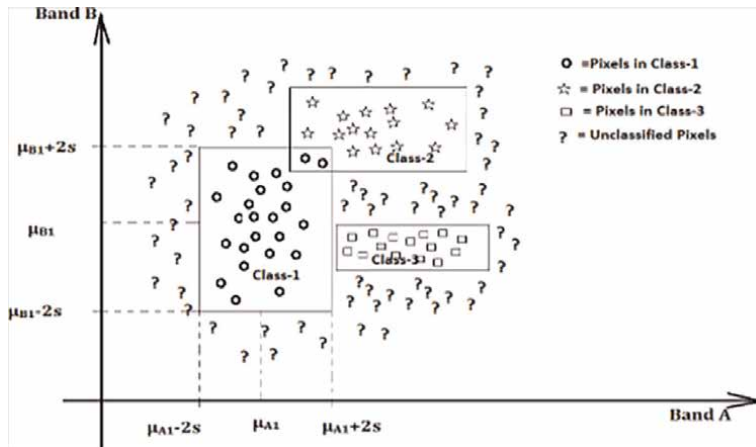


Figure 8. Classification of pixel data based on lowest ($\mu_{A1} - 2s$), highest ($\mu_{A1} + 2s$), and mean (μ_{A1}) values on band a and lowest ($\mu_{B1} - 2s$), highest ($\mu_{B1} + 2s$) and mean (μ_{B1}) values on band B of class-1 using Parallel Piped approach [16].

ranges for each class. However, a significant drawback is the possibility of overlapping boxes, which can introduce ambiguity in class assignment. These overlaps occur when the pre-established ranges for various classes intersect, making it challenging to definitively assign pixels in these overlapping areas to a single, distinct class.

2.7 Maximum Likelihood classification (MLC)

The Maximum Likelihood Classifier (MLC), also known as the Bayesian classifier, is a statistical approach used in supervised classification, as detailed in source [11]. This method classifies pixels into specific classes based on the maximum likelihood estimate that the pixels belong to those classes. The likelihood L_k of a pixel belonging to a particular class k is computed using its posterior probability, as expressed in Eq. (8). The primary goal of this technique is to find the class that maximizes this likelihood, effectively determining the most probable class for the given pixel. Due to its effectiveness in making class assignments grounded in probabilistic logic, the MLC is widely used in the field of remote sensing and various classification tasks.

$$L_k = \frac{P(k) * P\left(\frac{K}{x}\right)}{P(i) * P\left(\frac{K}{i}\right)} \quad (8)$$

where, $P(k)$ = prior probability of class k , $P(k/x)$ = conditional probability or probability density function of class K .

$P(k)$ and $P(i) * P(k/i)$ are common to all classes. So, L_k depends on the probability density function (**Figure 9**).

As demonstrated in Eq. (9), the probability density function (PDF) plays a pivotal role in computing the likelihood of a normal distribution, and its expression is presented as follows [17]. This function serves as a mathematical representation of the probability distribution of data points within a normal distribution, enabling the estimation of the likelihood associated with a specific observation's membership in this distribution.

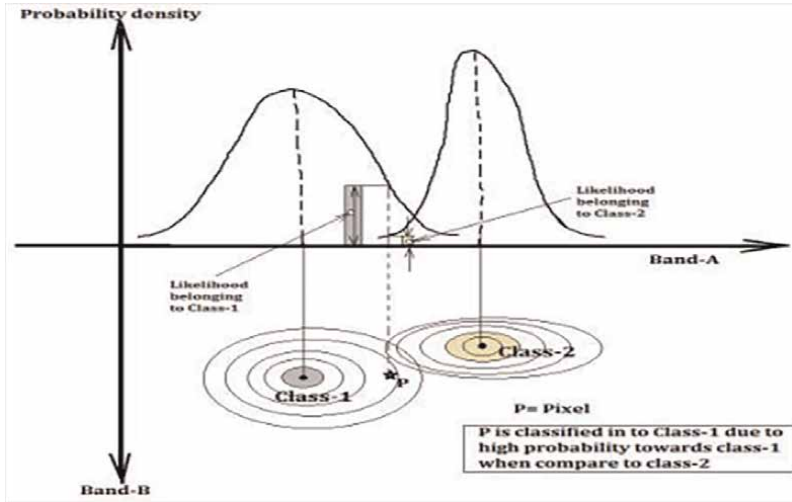


Figure 9.
Concept of Maximum Likelihood classification based on the probability density concerning each class [16].

$$L_k(X) = \frac{1}{\sqrt{(2\pi)^n} \gamma_k} e^{-\frac{(X - x_k)^2}{2 \gamma_k^2}} \quad (9)$$

$L_k(X)$ = likelihood of pixel X belonging to class, k , n = number of satellite image bands, and γ_k = variance-covariance matrix of class k .

2.8 K-Nearest Neighbor Classifier

As elucidated by [14], K-Nearest Neighbors (KNN) is a classification method that primarily relies on majority voting. The algorithm assesses similarity by computing the Euclidean distance between data points in the feature space, as depicted in Eq. (10). KNN employs a predefined parameter, K , to specify the number of nearest neighbors considered in the voting process. Unlike many other classifiers, KNN does not require a specialized training phase. It accomplishes classification by assigning a given data point to the most frequently occurring class among its K -nearest neighbors. This inherent simplicity renders KNN particularly valuable in scenarios where the training dataset may be subject to updates or expansions.

$$\begin{aligned} \text{Let } (x, y) \in D &\rightarrow \text{Set of training examples} \\ k &\rightarrow \text{Number of nearest neighbors} \\ z = (x', y') &\rightarrow \text{Test example} \end{aligned} \quad (10)$$

$$\text{Euclidean Distance } d(x', x) = \sum_{k=0}^n \left((x'_k - x_k)^2 \right)^{1/2}$$

In the K-Nearest Neighbors (KNN) algorithm, the classification of a test example is accomplished by identifying its K -nearest neighbors within the feature space, a calculation delineated by the method outlined in Eq. (11). Once the nearest neighbors are determined, the test example is assigned a label based on the majority class within this neighbor list. The underlying principle guiding this approach is that of spatial

proximity: the class that appears most frequently among the closest neighbors of the test example is considered the most probable class for the test example itself. This technique harnesses the strength of local information to make global predictions.

$$\text{Majority voting : } y' = \operatorname{argmax}_{(x_i, y_i) \in D_z} I(v = y_i)$$

(11)

Where, $v \rightarrow$ class label
 $y_i \rightarrow$ Class label for one of the nearest neighbors.

In certain scenarios, the K-Nearest Neighbors (KNN) classifier may encounter challenges when classifying certain test records. Such situations arise when the test records bear limited resemblance to or have little correspondence with the training instances. As KNN heavily relies on the proximity of test records to training data points, if there are no nearby neighbors from the training set within the specified value of K, the classifier may struggle to make confident classifications. This issue is particularly pronounced when dealing with sparse or high-dimensional datasets, where accurately defining the distances between data points becomes a complex task. To address this challenge, various techniques can be employed, including data preprocessing, feature selection, or dimensionality reduction, to enhance the performance of the KNN classifier and bolster its ability to effectively classify diverse test records.

2.9 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANNs) constitute a specialized subset of artificial intelligence designed to emulate certain aspects of the human brain, particularly the capability to assign meaningful labels to individual pixels in images. ANNs employ a nonparametric classification approach and possess inherent adaptability, allowing for the integration of additional data to enhance accuracy, as demonstrated in prior research [5]. Structurally, an ANN is organized into multiple layers, each containing neurons, which are processing units. These neurons are interconnected through weighted connections spanning across adjacent layers. During the training phase, ANNs analyze patterns within the training data to derive actionable insights. These insights aid in the formulation of generalizable rules that can be applied to classify previously unseen data, aligning with the foundational principles proposed by [22]. ANNs prove highly effective in addressing complex classification challenges, consistently outperforming more methods that are rudimentary. Numerous ANN algorithms are specifically tailored for classifying remotely sensed images, including but not limited to:

2.9.1 Multi-layer perceptron (MLP)

The Multi-layer perceptron (MLP) represents one of the most widely utilized subtypes of Artificial Neural Networks (ANNs). Operating as a feed-forward ANN model, the MLP is designed to establish mappings between sets of input data and their respective outputs, a concept that was initially introduced in prior research [23]. Structurally, the MLP comprises three essential layers: an input layer, an output layer, and one or more hidden layers. These layers are interconnected sequentially, as depicted in **Figure 2**. Users enjoy the flexibility to determine the number of neurons allocated to each layer. Specifically, the neurons in the input layer represent the input variables, while those in the output layer correspond to the categories defined by the training data. Typically, each input variable is mapped to a single neuron in the input

layer, and each class of interest is associated with its own neuron in the output layer. The training process of the MLP employs a supervised learning algorithm commonly known as back-propagation. The mathematical formalism for this process can be expressed as follows:

$$y = \varphi \left(\sum_{i=1}^n w_i x_i + b \right) = \varphi (w^T x + b) \quad (12)$$

where:

- W = represents the vector of weights.
- x = is the vector of inputs.
- b = stands for the bias.
- φ = denotes the activation function, typically chosen as the sigmoid function, which is expressed as $1/(1 + e^{-x})$. As demonstrated by [24], this particular activation function has proven effective in modeling nonlinear mappings.

In the framework of MLP, the weights and activation functions serve as critical elements within the neurons. During the training phase, neurons in both the input and hidden layers are initialized with random weights. Each pixel in the training data is then mapped to a likelihood of activating a specific output neuron, based on its maximum activation value. Subsequent solutions are evaluated in comparison, with the solution yielding the lowest error being retained. This iterative procedure persists until the network achieves an acceptable level of testing error for mapping input variables to the predetermined output classes. Once the network is adequately trained, it is deployed to classify the remaining dataset according to the activation level demonstrated by individual pixels, following the methodology described in [25]. However, one significant challenge associated with MLP is the necessity to retrain the entire network for any required adjustments. Such retraining could alter or nullify previously learned information, thereby prolonging the training duration—even when dealing with relatively modest datasets, as observed by [26]. To augment the efficiency of MLP classifiers without a corresponding substantial increase in computational time [22, 27], researchers have suggested specific parameter values, which are summarized in **Table 1**. Here, “ N ” denotes the number of classes (**Figure 10**).

Number of hidden layers/nodes	Between $2N_i$ to $3N_i$
Learning rate	0.01–0.001
Momentum factor	0.5
Sigmoid constant	1
RMSE	0.0001
Number of iterations	10,000
Accuracy rate percent	100%

Table 1.
The basic architecture to start MLP [28].

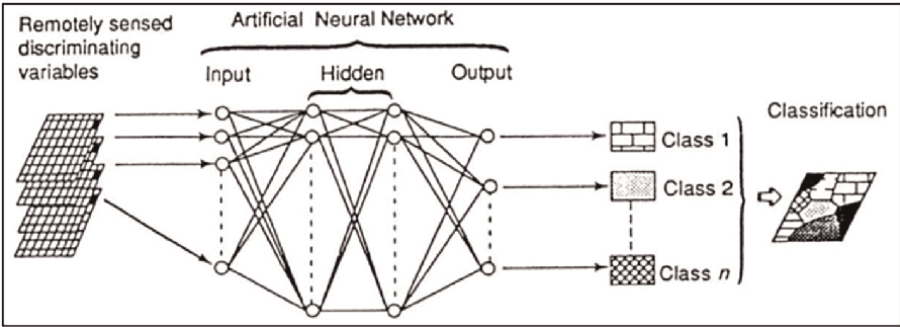


Figure 10.
A typical MLP with back-propagation [22] reported by [28].

2.9.2 Fuzzy ArtMap classification

Fuzzy ArtMap is a classification technique rooted in Adaptive Resonance Theory (ART), as originally introduced by [29]. This approach employs a clustering methodology designed for vectors with fuzzy inputs, which are represented as real numbers within the range of 0 to 1. Fuzzy ArtMap incorporates an incremental learning mechanism, allowing for continuous learning while preserving previously acquired knowledge, as elaborated upon by [30]. A distinctive feature of Fuzzy ArtMap is its emphasis on retaining only the weights of neurons encoding the class that best matches the input pattern. This characteristic enables it to effectively handle large-scale problems with relatively few training epochs. However, Fuzzy ArtMap is susceptible to noise and outliers, which can lead to an increased number of misclassified data points. The Fuzzy ArtMap architecture consists of four layers of neurons: the input layer (F1), category layer (F2), map field, and output layer. To configure Fuzzy ArtMap effectively, five key parameters must be specified, as outlined in **Table 2** by [31]:

The parameters ρ_1 and ρ_2 play a pivotal role in both the learning and operational phases of the network, exerting significant influence over the entire process. During the learning phase, the weights of neurons in both the map field and the category layer undergo dynamic adaptation. This adaptability is driven by the spectral characteristics inherent to each observed input pixel in the first input layer (F1). The assignment of an observation from the input layer (F1) to a neuron in the category layer (F2) hinges on the spectral features of the given observation. In instances where none of the existing neurons in F2 satisfies the predefined similarity threshold criteria for a particular observation in F1, a new F2 neuron is created. The generation of new neurons in F2 effectively partitions the dataset into subsets characterized by a user-specified

Choice parameter α	A small positive constant
Learning rate parameters β_1 in ARTa	$0 \leq \beta_1 \leq 1$
Learning rate parameters β_2 in ARTb	$0 \leq \beta_2 \leq 1$
Vigilance parameters ρ_1 in ARTa	Normally set very close to 1
Vigilance parameters ρ_2 in ARTb	Normally set very close to 1

Table 2.
The proposed parameters to start Fuzzy ArtMap [28].

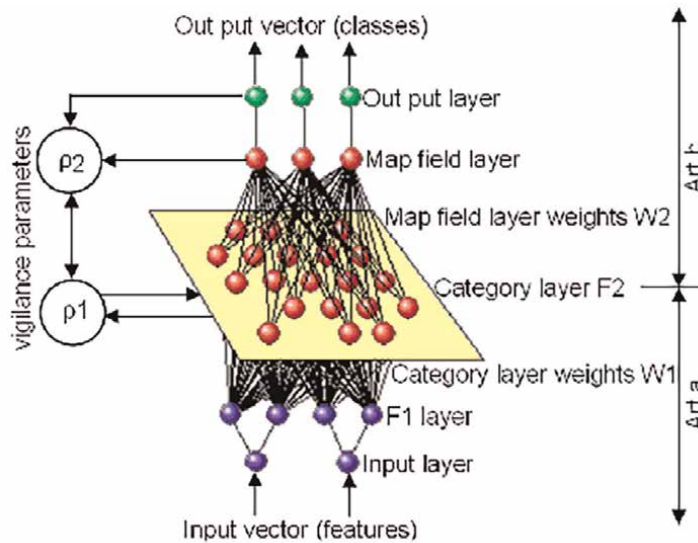


Figure 11.
Fuzzy ArtMap architecture [32] reported by [28].

degree of homogeneity (**Figure 11**). This degree of homogeneity is governed by a “vigilance parameter,” originally conceptualized by [15].

2.9.3 Self-Organized Feature Map (SOM)

SOM, initially proposed by [33], features a single-layer neural network architecture, as depicted in **Figure 4**. In this setup, the input layer corresponds to the feature vector, with individual neurons representing separate measurement dimensions. According to guidelines from [34], a typical output layer in an SOM is arranged as a 15 x 15 neuron grid. Arrays smaller than this may compromise the granularity needed for distinct class representation, while larger arrays generally improve classification accuracy. The synaptic weights linking the neurons of the output layer to the input layer neurons are initially randomized. These weights are subsequently reorganized through iterative sampling of the input data. This adjustment process leads to the spatial clustering of neurons with similar synaptic weights, as shown in **Figure 12**. During the training phase, neurons close to the ‘winning’ neuron—determined by the closeness between the input vector and its synaptic weight, as described in [35]—are actively engaged in learning. In this context, $X = (x_1, x_2, x_3, \dots, x_n)$ signifies the SOM’s reflectance vector for a single pixel input. The training process consists of two primary stages:

- a. Unsupervised classification phase: neuron weights are regionally structured through competitive learning and lateral interactions, thereby establishing the network’s topology.
- b. Refinement phase: the decision boundaries delineating various classes are further refined by integrating training samples via the Learning Vector Quantization (LVQ) algorithm, as elaborated by [36].

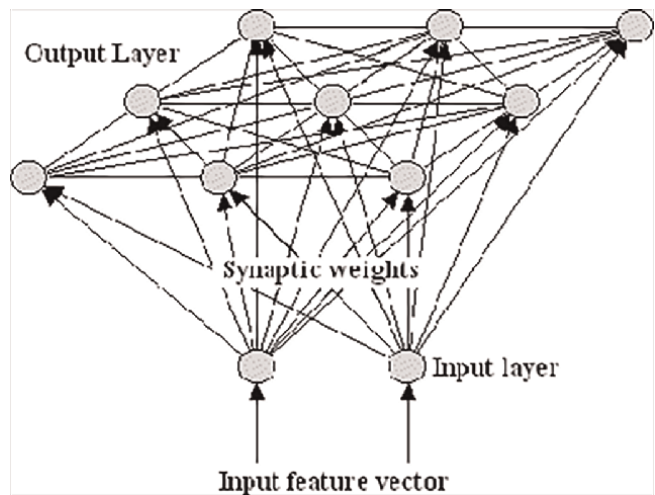


Figure 12.
Example of an SOM with a two-neuron input layer and 3×3 -neuron output layer [28].

In the classification step, each pixel is assigned a class based on the neuron or neurons with weight vectors most similar (in terms of minimal Euclidean distance) to the pixel's reflectance vector. Unlike other architectures like MLP or Fuzzy ArtMap, SOM takes into account inter-class relationships, facilitating the classification of multimodal classes. A notable drawback, however, is the occasional emergence of unclassified pixels, as highlighted by [37]. To optimize classification accuracy without substantially increasing computational load, recommended parameter settings for SOM classifiers are presented in **Table 3**.

2.9.4 Radial Basis Function Network (RBFN)

The Radial Basis Function Network (RBFN) is a type of nonlinear neural network designed for classification tasks. This network architecture consists of three primary layers: an n -dimensional input vector, a hidden layer of Radial Basis Function (RBF) neurons, and an output layer with individual nodes corresponding to each data category or class. The central aim of an RBFN is to classify data by evaluating the similarity between incoming inputs and stored prototypes within the RBF layer. Each neuron in the RBF layer contains a prototype, which is essentially a representative example from the training dataset. A commonly used method for selecting these prototypes involves applying k -means clustering to the training data, with the resultant cluster centers serving as the prototypes. During the network's operation, each RBF neuron calculates the Euclidean distance between its stored prototype and the incoming input vector. This distance is then converted into an activation value that falls between 0

Course tuning parameters				Fine-tuning parameters		
Output neurons	Initial	Min.	Max.	Min.	Max.	Fine-tuning
	γ	α	α	δ^t	δ^t	epoch
225 (15×15)	25	0.5	1	0.0001	0.0005	50

Table 3.
The proposed parameters to start SOM [28].

and 1, serving as an indicator of similarity. In this context, an activation value of 1 signifies a perfect match between the input and the prototype, while the activation value decreases exponentially toward zero as the distance between the input and the prototype grows. The output layer's nodes then compute a score for each potential data category. This score is obtained through a weighted sum of the activation values generated by the RBF neurons in the hidden layer. Before contributing to the final score, each neuron's activation is multiplied by its corresponding weight. The final classification decision is usually made by allocating the input to the category associated with the highest computed score (**Figure 13**).

Various similarity functions can be employed in the context of Radial Basis Function Networks (RBFNs), among which the Gaussian function is often favored, particularly for one-dimensional data. For instance, Eq. (13) outlines the form of a typical Gaussian function where 'x' is the input, 'μ' represents the mean, and 'σ' stands for the standard deviation. Notably, the activation function for an RBF neuron does exhibit slight variations, as shown in Eq. (14). The training regimen for an RBFN requires the optimization of three key parameter sets: the prototypes (μ), the β coefficients associated with each RBF neuron, and the output weight matrix connecting the RBF neurons to the nodes in the output layer. To improve classification accuracy, [38] suggests setting the μ and β parameters at specific values, namely 1.05 and 5, respectively.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (13)$$

$$\varphi(x) = e^{-\beta \|x-\mu\|^2} \quad (14)$$

2.10 Classification Trees (CTs)

Classification Trees (CTs) were originally conceptualized by [39] as a non-parametric, iterative, and hierarchical approach to pattern recognition. A Classification Tree consists of multiple essential components: a root node, which serves as the starting point; non-terminal nodes, which are positioned between the root and terminal nodes; and terminal nodes, which group pixels belonging to the same class, as

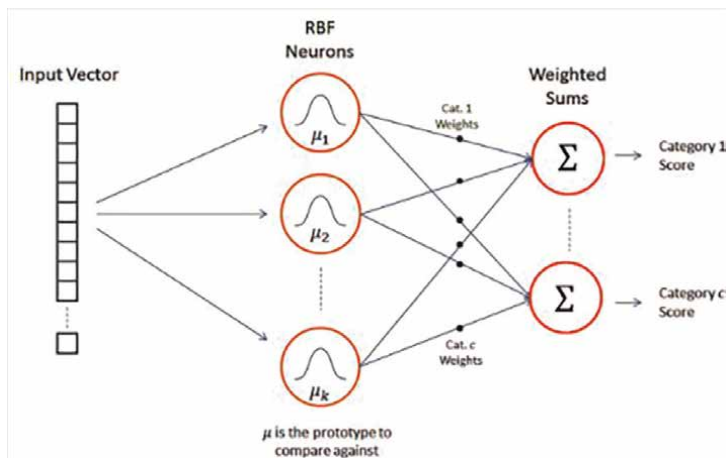


Figure 13.
RBF network architecture [28].

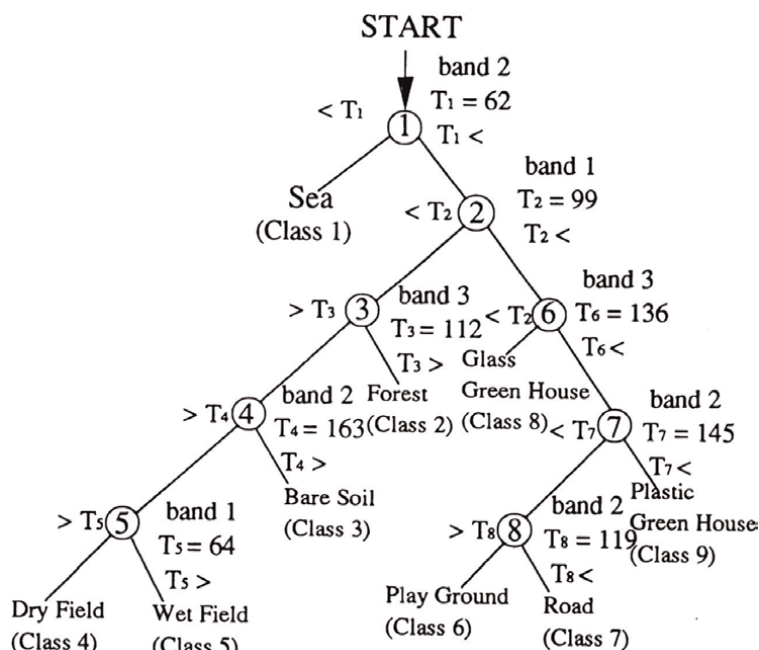


Figure 14.
 Classification tree. The numbers indicate the variables and their values as thresholds for each node condition [28].

illustrated in **Figure 14**. One of the key strengths of CTs is their capacity to recursively divide a dataset into increasingly homogeneous subsets, employing binary splitting rules as elaborated by [15]. These splitting rules are informed by the training data and are based on statistical techniques, with the metric of ‘impurity’ playing a crucial role. A node is considered ‘pure’ when it contains pixels exclusively from a single category, leading to an impurity score of zero. As the algorithm traverses the tree, it evaluates a logical condition at each node. Depending on the outcome of this evaluation, it moves to either the left or the right child node. This iterative partitioning persists until a node reaches a state of purity, at which point it is labeled as a terminal node.

The key splitting rules commonly used in Classification Trees (CTs) include:

1. *Entropy*: originated by [40], this measure assesses node homogeneity and aims to minimize entropy, ideally driving it to zero for a pure terminal node.
2. *Information Gain (IG) or Gain ratio*: further developed by [41], IG calculates the entropy reduction resulting from splitting a node N according to rule T. The attribute offering the highest IG is chosen for the split. One limitation is its bias toward variables with greater variances, which may lead to excessive splitting. This is often mitigated by normalizing the Information Gain with respect to the entropy of the created partitions.
3. *Gini index*: based on methods described by [39], this index measures node impurity and aims to find the most homogenous subsets within the training data. The rule that maximizes impurity reduction—or equivalently minimizes the Gini index—is selected for the split.

The potential for overfitting exists when a CT becomes too tailored to the training data, capturing noise that may affect its generalization capability. Pruning strategies are employed to counter this, often reducing the tree size by up to 25% and enhancing its predictive performance. Among the pruning methods, 10-fold cross-validation has gained recognition for its robustness, necessitating an independent dataset for performance assessment. Typically, pruned trees deliver the best classification accuracy. For a more comprehensive understanding of cross-validation, see [42].

2.11 Fuzzy classifiers

Fuzzy classifiers quantify the degree of fuzzy set membership for each pixel concerning specific classes. This degree is determined by calculating the standardized Euclidean distance from the mean of the class signature through a particular algorithm. In essence, the mean of a class signature symbolizes the optimal point for that class, where fuzzy set membership attains its maximum value of 1. As the distance from this point increases, the membership value diminishes, reaching zero at a user-defined distance threshold. Among the myriad fuzzy classification algorithms, Fuzzy C-Means (FCM) stands out for its effectiveness, especially in tasks involving overlapping clusters. Initially introduced by [43], FCM aims to minimize a fuzzy objective function, denoted as J_m , in an iterative optimization process, as illustrated in Eq. (15).

$$J_m = \sum_{i=1}^c \sum_{k=1}^n (\mu_{ik})^m d^2(x_k, V_i) \quad (15)$$

where:

- c = number of clusters
- n = total pixel count
- μ_{ik} = membership value for the i -th cluster of the k -th pixel
- m = fuzziness factor for each fuzzy membership
- x_k = vector representing the k -th pixel
- V_i = centroid vector of the i -th cluster
- $2d^2(x_k, V_i)$ = squared Euclidean distance between x_k and V_i

The membership (μ_{ik}) is calculated by measuring the distance between the k -th pixel's vector and the centroid of the i -th cluster. These calculated membership values are subject to the following constraints:

$$\left\{ \begin{array}{ll} 0 \leq \mu_{ik} \leq 1 & \text{for all } i, k \\ \sum_{i=1}^c \mu_{ik} = 1 & \text{for all } k \\ 0 < \sum_{k=1}^n \mu_{ik} < n & \text{for all } i \end{array} \right. \quad (16)$$

The cluster center (V_i) and the membership value (μ_{ik}) can be calculated using Eqs. (17) and (18), respectively.

$$V_i = \frac{\sum_{k=1}^n (\mu_{ik})^m x_k}{\sum_{k=1}^n (\mu_{ik})^m}, 1 \leq i \leq c \quad (17)$$

$$\mu_{ik} = \left[\sum_{j=1}^c \left(\frac{d(x_k, V_i)}{d(x_k, V_j)} \right)^{\frac{2}{m-1}} \right]^{-1}, 1 \leq i \leq c, 1 \leq k \leq n \quad (18)$$

To minimize J_m , a repetitive procedure is followed, governed by Eqs. (17) and (18). The process begins by establishing fixed values for c (the number of clusters), the degree of fuzziness m , and a convergence tolerance ϵ , in addition to initializing the starting centers for each cluster. Subsequently, the values of μ_{ik} and V_i are computed in accordance with Eqs. (17) and (18). The iterative process continues until the variation in V_i between two consecutive iterations falls below the specified threshold ϵ . Eventually, each pixel is assigned a category based on its varying degrees of membership across multiple clusters.

3. Hybrid Classifiers

Different classifiers often yield varying class assignments when applied to the same test area, highlighting the fact that no single classifier can deliver optimal performance across all classes. Various classification algorithms exhibit complementary characteristics, as supported by the research presented in [44]. Neural and statistical algorithms, in particular, demonstrate such complementary behavior. These classifiers tend to produce uncorrelated classification errors, which in turn facilitate the attainment of improved classification accuracies when they are combined. In the context of a hybrid classifier-based approach, it is crucial for the individual classifiers to either employ independent sets of features or be trained on distinct training datasets. Essentially, there are two primary strategies for combining classifiers:

1. Classifier ensembles (CE).
2. Multiple classifier systems (MCS), as illustrated in **Figure 15**.

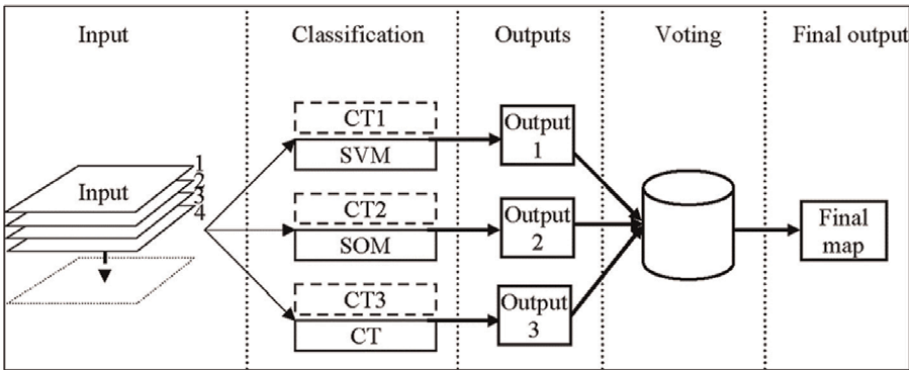


Figure 15.
Classifier ensemble (dashed) versus multiple classifier systems (solid) [28].

In situations where the classification outcomes from multiple classifiers demonstrate a high degree of similarity, the amalgamation of these classifiers may only sporadically yield improvements in classification accuracy. This underscores the pivotal importance of classifier diversity for the effective implementation of hybrid systems, a point emphasized by the findings reported in [45]. Although not frequently employed or scrutinized in the context of remote-sensing image classification, measures of classifier diversity are indispensable for evaluating the variations among different algorithms. Such metrics encompass Kappa statistics [46], the double fault measure [47], agreement measures [48], similarity indices, measures of non-inferiority and difference [49], weighted counts of errors and correct results (WCEC) [50], entropy-based metrics [51], and the Disagreement-accuracy measure [52].

4. Comparison of satellite image classification methods

Numerous studies have been undertaken to compare unsupervised, supervised, and hybrid methods for satellite image classification, focusing on evaluation criteria such as classification accuracy and various alphanumeric constants. This section offers a comprehensive review of these comparative analyses. **Table 4** delineates the merits and demerits of conventional classifiers, as cited in [53], while **Table 5** engages in a more detailed discussion concerning the advantages and disadvantages inherent to various classification approaches. The task of selecting the most apt classifier for a given dataset poses a significant challenge, chiefly due to the absence of straightforward guidelines to facilitate this decision-making process. Furthermore, the image

Classifier	Advantages	Disadvantages
ISODATA	• Fast and simple to process	• Needs several parameters
K-means	• Fast and simple to process	• Could be influenced by: the number and position of the initial cluster centers specified by the analyst, the geometric properties of the data, and clustering parameters
KNN	• Simple to process	• Computationally expensive when the training dataset is large
MD	• Fast and simple to process	• Considers only mean value
MhD	• Fast and simple to process	• Considers only mean value
PP	• Fast and simple to process	• Overlap may reduce the accuracy of the results
MXL	• Sub-pixel classifier	• Time consuming • Insufficient ground truth data and/or correlated bands can affect the results • Cannot be applied when the dataset is not normally distributed

Table 4.
Comparison of classic classification techniques.

Method	Advantages	Disadvantages
ANN	• Non-parametric classifiers • High computation rate of very large datasets • Efficiently handles noisy inputs	• It is difficult to understand how the result was achieved. • The training process is slow. • Problem of over fitting • Difficult to select the type network architecture • Dependent on user-defined parameters
CTs	• Non-parametric classifiers • Does not require an extensive design and training • Easy to understand the classification process • Accurate, and computational efficiency is good • Easy to integrate multi-source data	• Calculation becomes complex when various outcomes are correlated.
SVMs	• Non-parametric classifiers • Provides a good generalization • The problem of over fitting is controlled. • Computational efficiency is good. • Perform well with minimum training set size and high-dimensional data • Often outperform other classifiers.	• Training is time consuming. • Difficult to understand its structure • Dependent on user-defined parameters • Determination of optimal parameters is not easy.
Fuzzy Classifiers	• Efficiently handle overlapping data • Minimize computation time and reduce memory requirements	• Without priori knowledge output is not good.

Table 5.
Comparison of modern classification techniques [54].

Researcher	Classifier	Datasets	Best performance
Pal and Mather [55]	• SVMs • MXL • ANN	• Landsat 7 ETM+ • Hyperspectral data	SVMs
Oliveira et al. [30]	• MXL • CTs • MLP • SOM • ArtMap • CTs	• Landsat (ETM+)	ArtMap
Lippitt et al. [56]	• SOM • MLP • ArtMap • CTs	• Landsat 7 (ETM+)	CTs
Li et al. [31]	• MXL • CTA • ArtMap • KNN	• Landsat 5 TM • ALSO PALSAR (L-band HH and HV)	ArtMap
Maryam et al. [57]	• SVMs • MXL • MD • PP	• Landsat 7 ETM+	SVMs

Researcher	Classifier	Datasets	Best performance
Shaker et al. [58]	- Contextual - MXL - MD	SPOT	MXL
Mannan et al. [59]	- ArtMap - MXL - MLP	IRS-1B	ArtMap
Gil et al. [60]	- SVMs - ANN - MXL	IKONS	SVMs
Du et al. [52]	- MLP - ANN - CTs - MXL - SVMs - ArtMap	- Quick Bird - Spectral/textural features	SVMs
Doma et al. [61]	- PP - MD - MXL - ANN - SVMs	Quick Bird	SVMs
Hamedianfar et al. [62]	- OO/fuzzy - SVMs	World View-2 (WV-2)	OO/fuzzy
Camps-Valls et al. [63]	- SVMs - ANN	Hyperspectral data (128 bands)	SVMs
Trinder et al. [64]	- SVMs - SOM - CTs	- Aerial images - LiDAR data	SVMs
Kanika et al. [65]	- ANN - MD - MLC	IRIS Plants Dataset	ANN
Jamshid et al. [66]	- ISODATA - MLC - HM	Desert Outlay Datasets	HM
Offer and Arnon [67]	- MLC - MD - PP	Landsat 7 ETM+	MLC
Aykut et al. [68]	- PP - MD - CM	Landsat 5 TM	CM
Shila and Ali [69]	- Unsupervised - Supervised - HM	Landsat 7 ETM +	HM
Maryam et al. [57]	- SVMs - MLC - MHD - MD - SID - BCs - PP	Landsat 7 ETM +	SVMs

Researcher	Classifier	Datasets	Best performance
Manoj et al. [14]	• KM • ISODATA • MD • MLC • PP • SRG • ESRG	Landsat, SPOT and IRS Datasets	ESRG
Subhash et al. [70]	• MLC • MD • MHD	Landsat 7 ETM +	MLC
Malgorzata et al. [71]	• PBC • OOC	Multi-spectral Satellite Images	OOC

Table 6.
Performance evaluation of various classification methods against different datasets.

classification workflow involves a series of decisions and choices that must be made by the analyst. To elucidate the comparative strengths and limitations of unsupervised, supervised, and hybrid techniques, extensive comparative studies have been conducted. **Table 6** synthesizes findings from various researchers concerning the efficacy of different classifiers and the contexts in which they are most effective. It is noteworthy that there exists no universal agreement among researchers as to which classification methodology is unequivocally superior. As a result, the choice of an appropriate classifier is often contingent upon the specific requirements and objectives of the project at hand.

5. Selection and evaluation process of satellite image classification methods

The evaluation process and the selection of methods and techniques for satellite image classification represent a critical phase in any remote-sensing or Earth observation study. This stage involves careful consideration of various factors, including the characteristics of the satellite data, the specific objectives of the classification task, and the computational resources available. Researchers must weigh the strengths and weaknesses of different classification algorithms, such as Artificial Neural Networks, Support Vector Machines, and Classification Trees, in order to make an informed choice. Furthermore, evaluation criteria extend beyond mere accuracy and encompass factors like computational efficiency, ease of implementation, and the ability to handle challenges like overfitting or underfitting. This rigorous process ensures that the selected methods and techniques align with the project's objectives, providing meaningful insights into the Earth's dynamics and environmental changes based on satellite imagery, as described in **Figure 16**.

This figure provides a comprehensive overview of the systematic process employed in the selection and evaluation of satellite image classification methods and techniques. It guides readers through the intricate steps involved in methodological decision-making, offering a clear depiction of how various algorithms are chosen, assessed, and ultimately applied to satellite image classification tasks. This visual representation enhances the clarity of our research, helping both novice and

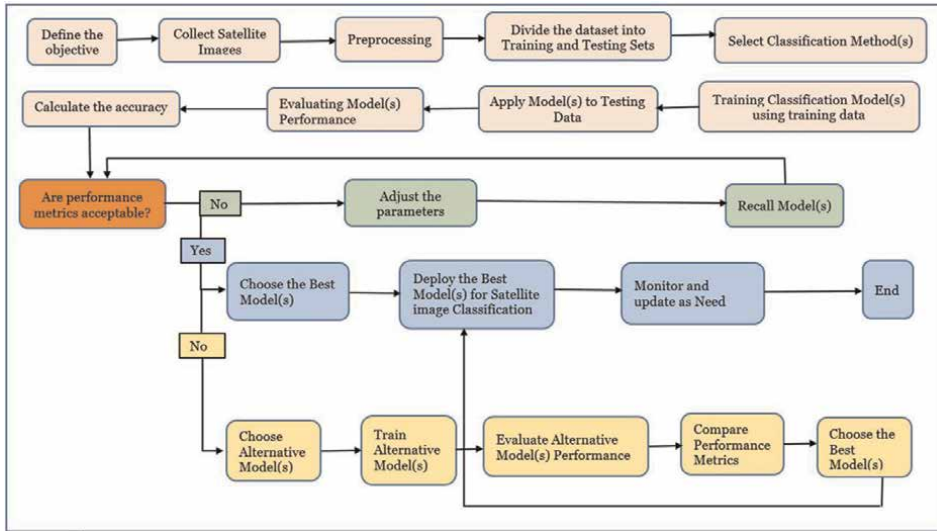


Figure 16.
Selection and evaluation process of satellite image classification methods and techniques.

experienced researchers in the field to grasp the complexity of the decision-making process and its significance in achieving accurate and meaningful results.

6. Summary and discussion

In the classification of remotely sensed images, numerous Artificial Neural Network (ANN) algorithms have been implemented. Multilayer Perceptron (MLP), Self-Organizing Maps (SOM), and Fuzzy ArtMap are some examples. Fuzzy ArtMap has emerged as the most effective, followed closely by MLP. In contrast, SOM has consistently demonstrated the lowest classification accuracy across numerous investigations. Importantly, the performance of these algorithms is extremely dependent on the expertise of the operator in fine-tuning algorithmic parameters. MLP is effective, but it requires a complete reprogramming of the network, which is a disadvantage. This can be a time-consuming operation, particularly for smaller test areas. To mitigate this, pre-trained models could be used as a starting point. Fuzzy ArtMap, on the other hand, is adept at dealing with large-scale problems and requires fewer training iterations. However, it is susceptible to noise and outliers, which may impact classification accuracy negatively. Preprocessing steps to remove noise can help overcome this challenge. SOM, unlike MLP and Fuzzy ArtMap, is capable of distinguishing between multimodal classes but frequently leaves a greater number of pixels unclassified. This can be mitigated by adjusting parameters or using hybrid approaches. Entropy splitting has become the algorithm of preference when it comes to Classification Trees (CTs) for image classification. The 10-fold cross-validation method is also recognized for its dependability and precision. CTs developed for one test area can frequently be successfully transferred to another, provided that the remotely sensed images share similar sensor characteristics and Land Use and Land Cover (LULC) patterns. Support Vector Machines (SVMs) flourish in classification precision and class-specific error balance. The Radial Basis Function (RBF) kernel is frequently

the best option among the numerous kernels. However, determining the optimal RBF kernel parameters, specifically C , typically requires a grid search and 10-fold cross-validation. It is essential to recognize that various classifiers provide distinct data perspectives. While some classifiers excel at identifying particular classes, others are better adapted for distinguishing between distinct groups. Consequently, combining classifiers efficiently can result in greater classification accuracy than using a single classifier, even if it is extremely effective. When combined, neural and statistical classifiers frequently generate uncorrelated errors, thereby enhancing the overall classification accuracy. Adding more classifiers does not necessarily enhance performance, though. Diversity is a crucial success factor for hybrid systems. Combinations of classifiers selected based on the Disagreement accuracy metric typically outperform those selected using other diversity metrics. Both Classifier Ensembles (CE) and Multiple Classifier Systems (MCS) are based on aggregating multiple instances of the identical algorithm.

7. Conclusion

In recent decades, significant progress has been made in the field of picture classification, mostly due to the emergence and utilization of sophisticated classification algorithms. This comprehensive review aims to provide invaluable guidance for researchers by dissecting the strengths and limitations of each approach, with special emphasis on their suitability in various contexts. Artificial Neural Networks, particularly Multilayer Perceptrons (MLPs) and Fuzzy ArtMap, present distinct considerations, chiefly revolving around computational complexity and parameter tuning. MLPs often necessitate an extensive network reconfiguration, a time-consuming process, especially for smaller datasets. Nevertheless, their adaptability makes them well-suited for intricate classification tasks. Fuzzy ArtMap excels in managing large-scale problems with fewer training iterations but is sensitive to noise and outliers, demanding meticulous pre-processing for accuracy. In the case of Support Vector Machines, the selection of a kernel function and its associated parameters significantly influence classification outcomes. The Radial Basis Function (RBF) kernel frequently outperforms others in terms of effectiveness, yet pinpointing optimal parameters typically involves computationally intensive grid search and 10-fold cross-validation. SVMs excel at maintaining class-specific error balance but may encounter challenges with excessively large or imbalanced datasets. On the contrary, Classification Trees offer interpretability and ease of use as their primary advantages. They possess the ability to conduct feature selection as an integral part of the model-building process, effectively reducing dimensionality. However, they are susceptible to overfitting, particularly when the tree structure becomes overly complex. Techniques such as pruning and cross-validation can help alleviate this concern. While entropy-based splitting rules have proven effective, the algorithm's performance may vary depending on the chosen impurity measure. Additionally, this review underscores the vital role of algorithmic optimization, elaborating on the selection and combination of base classifiers to enhance performance. Diversity measures assume a pivotal role in guiding ensemble construction, ensuring more reliable and robust classification results. The review synthesizes these intricate details by drawing upon collective insights from existing studies, enabling researchers to align their methodological preferences with project-specific requirements and constraints. Evaluation criteria encompass not only accuracy but also extend to considerations such as computational cost, ease of

implementation, and resilience to overfitting or underfitting, all tailored to the characteristics of each algorithm. This exhaustive study not only aggregates diverse viewpoints but also serves as a pragmatic roadmap for future research endeavors. By comprehensively presenting the advantages and practical challenges associated with each method, along with their specific evaluation criteria, it aims to empower researchers with a nuanced understanding, enabling them to make well-informed decisions within the realm of satellite image classification.

Acknowledgements

This chapter was supported by the project TKP2021-NVA-13 (BorderEye) which has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NVA funding scheme.

Author details

Emad H.E. Yasin^{1,2*} and Czimber Kornel²

1 Faculty of Forestry, Department of Forest Management, University of Khartoum, Khartoum North, Sudan

2 Institute of Geomatics and Civil Engineering, Faculty of Forestry, University of Sopron, Sopron, Hungary

*Address all correspondence to: emad.yasin823@gmail.com

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Jensen JR. Introductory Digital Image Processing: A Remote Sensing Perspective. 3rd ed. Upper Saddle River: Prentice-Hall; 2005
- [2] Chaichoke V, Supawee P, Tanasak V, Andrew KS. A normalized difference vegetation index (NDVI) time-series of idle agriculture lands: A preliminary study. *Engineering Journal*. 2011; **15**(1):9-16
- [3] Lu D, Weng Q. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*. 2007; **28**(5):823-870
- [4] Kumar S, Kumar D. A review of remotely sensed satellite image classification. *International Journal of Electrical and Computer Engineering*. 2019; **9**(3):1720
- [5] Abburu S, Golla S. Satellite image classification methods and techniques: A review. *International Journal of Computer Applications*. 2015; **119**(8):20-25
- [6] Kulkarni AD, Kamlesh L. Fuzzy neural network models for supervised classification: Multispectral image analysis. *Geocarto International*. 1999; **4**:42-51. DOI: 10.1080/10106049908542127
- [7] Wu C, Murray AT. Estimating impervious surface distribution by spectral mixture analysis. *Remote Sensing of Environment*. 2003; **84**: 493-505. DOI: 10.1016/S0034-4257(02)00136-0
- [8] Yang CC, Prasher SO, Enright P, Madramootoo C, Burgess M, Goel PK, et al. Application of decision tree technology for image classification using remote sensing data. *Agricultural Systems*. 2003; **76**:1101-1117. DOI: 10.1016/S0308-521X(02)00051-3
- [9] Blaschke T. Object based image analysis for remote sensing. *Journal of Photogrammetry and Remote Sensing (ISPRS)*. 2009; **65**:2-16. DOI: 10.1016/j.isprsjprs.2009.06.004
- [10] Pal NR, Bhandari D. On object background classification. *International Journal of Systems Science*. 1992; **23**: 1903-1920. DOI: 10.1080/00207729208949429
- [11] Benz UC, Hofmann P, Willhauck G, Lingenfelder I, Heynen M. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *Journal of Photogrammetry and Remote Sensing (ISPRS)*. 2004; **58**:239-258. DOI: 10.1016/j.isprsjprs.2003.10.002
- [12] Ahmed R, Mourad Z, Ahmed BH, Mohamed B. An optimal unsupervised satellite image segmentation approach based on Pearson system and k-means clustering algorithm initialization. *International Science Index*. 2016; **3**(11): 948-955
- [13] Al-Ahmadi FS, Hames AS. Comparison of four classification methods to extract land use and land cover from raw satellite images for some remote arid areas, Kingdom of Saudi Arabia. *Journal of King Abdul-Aziz University, Earth Sciences*. 2009; **20**(1): 167-191
- [14] Manoj P, Astha B, Potdar MB, Kalubarme MH, Bijendra A. Comparison of various classification techniques for satellite data. *International Journal of Scientific & Engineering Research*. 2013; **4**(2):1-6

- [15] Tso B, Mather PM. Classification Methods for Remotely Sensed Data. 2nd ed. America: Taylor and Francis Group; 2009
- [16] Rajyalakshmi D, Raju KK, Varma GS. Taxonomy of satellite image and validation using statistical inference. In: 2016 IEEE 6th International Conference on Advanced Computing (IACC). Bhimavaram, India: IEEE; 2016. pp. 352-361
- [17] Shalaby A, Tateishi R. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the northwestern coastal zone of Egypt. *Applied Geography*. 2006;27: 28-41. DOI: 10.1016/j.apgeog.2006.09.004
- [18] Marconcini M, Camps-Valls G, Bruzzone L. A composite semi-supervised SVM for classification of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*. 2009;6:234-238. DOI: 10.1109/LGRS.2008.2009324
- [19] Deer PJ, Eklund P. Study of parameter values for a Mahalanobis distance fuzzy classifier. *Fuzzy Sets and Systems*. 2003;137:191-213. DOI: 10.1016/S0165-0114(02)00220-8
- [20] Zhang B, Li S, Jia X, Gao L, Peng M. Adaptive Markov random field approach for classification of hyperspectral imagery. *Geoscience and Remote Sensing, IEEE*. 2011;8:973-977. DOI: 10.1109/LGRS.2011.2145353
- [21] Zhu HW, Basir O. An adaptive fuzzy evidential nearest neighbor formulation for classifying remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*. 2005;43:1874-1889. DOI: 10.1109/TGRS.2005.848706
- [22] Foody GM. Image classification with a neural network: From completely crisp to fully-fuzzy situations. In: Atkinson PM, Tate NJ, editors. *Advances in Remote Sensing and GIS Analysis*. Chichester: Wiley & Son; 1999
- [23] Rosenblatt F. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Washington DC: Spartan Books; 1962. Available from: <http://catalog.hathitrust.org/Record/000203591>
- [24] Cybenko G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*. 1989;2(4):303-314
- [25] Foody GM. Land-cover classification by an artificial neural network with ancillary information. *International Journal of Geographical Information Systems*. 1995;9(5):527-542
- [26] Liu W, Seto K, Wu E, Gopal S, Woodcock C. ARTMMAP: A neural network approach to subpixel classification. *IEEE Transactions on Geoscience and Remote Sensing*. 2004; 42(9):1976-1983
- [27] Kavzoglu T, Mather PM. The use of back-propagating artificial neural networks in land cover classification. *International Journal of Remote Sensing*. 2003;24(3):4907-4938
- [28] Salah M. A survey of modern classification techniques in remote sensing for improved image classification. *Journal of Geomatics*. 2017;11(1):21
- [29] Carpenter GA, Crossberg S, Reynolds JH. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*. 1991; 4(5):565-588
- [30] Oliveira L, Oliveira T, Carvalho L, Lacerda W, Campos S, Martinhago A.

- Comparison of machine learning algorithms for mapping the phytophysiologicals of the Brazilian Cerrado. In: IX Brazilian Symposium on GeoInformatics; November 25-28, 2007. Campos do Jordão, Brazil: INPE; 2007. pp. 195-205
- [31] Li G, Lu D, Moran E, Sant'Anna S. Comparative analysis of classification algorithms and multiple sensor data for land use/land cover classification in the Brazilian Amazon. *Journal of Applied Remote Sensing*. 2012;**6**(1):11
- [32] Eastman JR. Idrisi Andes: Tutorial. Clark Labs: Clark University, Worcester; 2006
- [33] Kohonen T. The self-organizing map. *Proceedings of the IEEE*. 1990;**78**: 1464-1480
- [34] Hugo C, Capao L, Fernando B, Mario C. MERIS based land cover classification with self-organizing maps: Preliminary results. In: *Proceedings of the 2nd EARSeL SIG Workshop on Land Use & Land Cover* (Unpaginated CD-ROM); September 28-30, 2006. Bonn, Germany: Center for Remote Sensing of Land Surfaces; 2007
- [35] Jen-Hon L, Din-Chang T. Self-organizing feature map for multi-spectral spot land cover classification. In: *GIS development.net, AARS, ACRS 2000*. Singapore: Asian Conference on Remote Sensing (ACRS)/Asian Association on Remote Sensing (AARS); 2000
- [36] Nasrabadi NM, Feng Y. Vector quantization of images based upon the Kohonen self-organizing feature maps. In: *Proceedings of the IEEE International Conference on Neural Networks (ICNN-88)*; July 24-27, 1988; San Diego, California. Institute of Electrical and Electronics Engineers (IEEE); 1988. pp. 101-108
- [37] Qiu F, Jensen JR. Opening the black box of neural networks for remote sensing image classification. *International Journal of Remote Sensing*. 2004;**25**(9):1749-1768
- [38] Hwang YS, Bang SY. An efficient method to construct a radial basis function neural network classifier. *Neural Networks*. 1997;**10**(8):1495-1503
- [39] Breiman L, Friedman JH, Olshen RA, Stone CJ, editors. *Classification and Regression Trees*. New York: Chapman & Hall; 1984. 358 p
- [40] Shannon CE, editor. *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press; 1949
- [41] Quinlan JR. Simplifying decision trees. *International Journal of Man-Machine Studies*. 1987;**27**(3):227-248
- [42] Sherrod PH. DTREG Tutorial Home Page. Available from: <https://www.dtreg.com/methodology> [Accessed 7 September 2023]
- [43] Bezdec JC. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press; 1981
- [44] Kanellopoulos I, Wilkinson G, Roli F, Austin J, editors. *Neurocomputation in Remote Sensing Data Analysis*. Berlin: Springer; 1997
- [45] Chandra A, Yao X. Evolving hybrid ensembles of learning machines for better generalization. *Neurocomputing*. 2006;**69**(7-9):686-700
- [46] Congalton RG, Mead RA. A quantitative method to test for consistency and correctness in

photointerpretation. *Photogrammetric Engineering and Remote Sensing*. 1983; **49**(1):69-74

[47] Giacinto G, Roli F. Design of effective neural network ensembles for image classification. *Image and Vision Computing* (Elsevier). 2001;**19**(9-10): 697-705

[48] Michail P, Benediktsson JA, Ioannis K. The effect of classifier agreement on the accuracy of the combined classifier in decision-level fusion. *IEEE Transactions on Geoscience and Remote Sensing*. 2002;**39**(11):2539-2546

[49] Foody GM. Classification accuracy comparison: Hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. *Remote Sensing of Environment*. 2009;**113**(8):1658-1663

[50] Aksela M, Laaksonen J. Using diversity of errors for selecting members of a committee classifier. *Pattern Recognition*. 2006;**39**:608-623

[51] Kuncheva LI, Whitaker CJ. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*. 2003;**51**(2): 181-207

[52] Du P, Xia J, Zhang W, Tan K, Liu Y, Liu S. Multiple classifier system for remote sensing image classification: A review. *Sensors*. 2012;**12**(4):4764-4792

[53] Richards JA. *Remote Sensing Digital Image Analysis*. 5th ed. Berlin: Springer-Verlag; 2013

[54] Kamavisdar P, Saluja S, Agrawal S. A survey on image classification approaches and techniques. *International Journal of Advanced Research in Computer and Communication Engineering*. 2013;**2**(1): 1005-1008

[55] Pal M, Mather PM. Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*. 2005;**26**(5):1007-1011

[56] Lippitt C, Rogan J, Li Z, Eastman J, Jones T. Mapping selective logging in mixed deciduous forest: A comparison of machine learning algorithms. *Photogrammetric Engineering and Remote Sensing*. 2008;**74**(10):1201-1211

[57] Maryam N, Vahid MZ, Mehdi H. Comparing different classifications of satellite imagery in forest mapping (case study: Zagros forests in Iran). *International Research Journal of Applied and Basic Sciences*. 2014;**8**(7): 1407-1415

[58] Shaker A, Yan WY, El-Ashmawy N. Panchromatic satellite image classification for flood hazard assessment. *Journal of Applied Research and Technology*. 2012;**10**(x):902-911

[59] Mannan B, Roy J, Ray AK. Fuzzy ArtMap supervised classification of multi-spectral remotely-sensed images. *International Journal of Remote Sensing*. 1998;**19**(4):767-774

[60] Gil A, Yu Q, Lobo A, Lourenço P, Silva L, Calado H. Assessing the effectiveness of high-resolution satellite imagery for vegetation mapping in small islands protected areas. *Journal of Coastal Research*. 2011;**2011**(64):1663-1667

[61] Doma ML, Gomaa MS, Amer RA. Sensitivity of pixel-based classifiers to training sample size in case of high-resolution satellite imagery. *Journal of Geomatics*. 2015;**9**(2):53-58

[62] Hamedianfar A, Mohd Shafri HZ, Mansor S, Ahmad N. Detailed urban object-based classifications from WorldView-2 imagery and LiDAR data:

Supervised vs. fuzzy rule-based. In: FIG Congress 2014, Engaging the Challenges—Enhancing the Relevance. Kuala Lumpur, Malaysia: XXV International Federation of Surveyors Congress; 16-21 June 2014

[63] Camps-Valls G, Gomez-Chova L, Calpe-Maravilla J, Soria-Olivas E, Martín-Guerrero JD, Moreno J. Support vector machines for crop classification using hyperspectral data. In: Pattern Recognition and Image Analysis: First Iberian Conference, IbPRIA 2003, Puerto de Andratx, Mallorca, Spain, June 4-6, 2003. Proceedings 1. Berlin, Heidelberg: Springer; 2003. pp. 134-141

[64] Trinder J, Salah M, Shaker A, Hamed M, Elsagheer A. Combining statistical and neural classifiers using Dempster-Shafer theory of evidence for improved building detection. In: Proceeding of 15th Australas. Remote Sens. Photogramm. Conf., Alice Springs, Australia. 13-17 September 2010. pp. 13-16

[65] Kanika K, Anil KG, Rhythm G. A comparative study of supervised image classification algorithms for satellite images. International Journal of Electrical, Electronics and Data Communication. 2013;1(10):10-16

[66] Jamshid T, Nasser L, Mina F. Satellite Image Classification Methods and Landsat 5tm Bands. Ithaca, New York: Cornell University Library; 2013

[67] Offer R, Arnon K. Comparison of methods for land-use classification incorporating remote sensing and GIS inputs. EARSel eProceedings. 2011;10(1):27-45

[68] Aykut A, Eronat AH, Necdet T. Comparing different satellite image classification methods: An application in Ayvalik District, Western Turkey. In: Proceedings of the XXth ISPRS Congress

Technical Commission, ISPRS, Vol. XXXV Part B4; July 12-23. Vol. 40. Istanbul, Turkey: The 4th International Congress for Photogrammetry and Remote Sensing; Jul 2004

[69] Shila HN, Ali RS. Comparison of land covers classification methods in ETM+ satellite images (case study: Ghamishloo wildlife refuge). Journal of Environmental Research and Development. 2010;5(2):279-293

[70] Subhash T, Akhilesh S, Seema S. Comparison of different image classification techniques for land use land cover classification: An application in Jabalpur District of Central India. International Journal of Remote Sensing and GIS. 2012;1(1):26-31

[71] Malgorzata VW, Anikó K, Éva V. Comparison of different image classification methods in urban environment. In: Proc. International Scientific Conference on Sustainable Development & Ecological Footprint; March 26-27, 2012. Sopron, Hungary: University of Sopron Press

Programming an Interactive Map to Enhance Clients' Participation in Site Selection of E-Marketing in Cairo and Giza Governorates: Egypt

Reda Alkot Mohamed and Sayed Ahmed Sayed

Abstract

Multiple-Criteria Decision Analysis (MCDA) is concerned with structuring and taking decisions involving multiple criteria. Integrated with Geographic Information System (GIS) technology, MCDA can be used to select the best locations for any investment project. However, not all investors/clients have the required skills to use these specialised methods and programmes. Also, there may be incompatibility between the consultant and the end user in the end results. Therefore, this research aims to program a web-based Decision Support Engine (DSE) to facilitate the decision-making process and involve the end user in such process in an easy interactive manner. The proposed DSE was applied on E-marketing applied study in Cairo and Giza governorates. Data were collected by electronic questionnaire (928 Questionnaires) that was processed by SPSS, to create the geodatabase used by Django web framework to build the DSE. The end result of DSE is producing an Interactive Map to support clients' decision-making in an easy manner.

Keywords: E-marketing, Cairo and Giza, interactive map, Django framework, DSE

1. Introduction

Feasibility studies are one of the most important studies that economic investors need on an ongoing basis in order to establish their economic projects because they intensely help them in making their decisions in the field of investment. One of the most important elements of the feasibility studies is to create a suitability map for the best locations to carry out these projects by the integration with GIS technology. GIS produces this map by applying a site selection model to determine the most suitable locations for agriculture, industry, trade, or other economic and life activities.

Also, GIS provides MCDA tools in spatial decision support and creating spatial modelling to support the proper decision in selecting the best location [1]. Site selection modelling process is applied for selecting the best locations for stores, shops, and the shortest paths. So, Geographic Decision Support (GDS) is a promising field in managing economic resources in general and for decision support efforts in the field

of urban planning and Spatial Decision support systems (SDSS) in specific [2]. Also, the land suitability model is applied through MCDA because the spatial decision-making process is a complex process, as it includes many and varied data (economic, social, environmental, political, and other data). Therefore, the decision-making process is called Multi-Criteria Decision-Making (MCDM).

But end users/clients may not have the required knowledge for using these specialised methods and programmes such as GIS. So, these experiences can be gained by sharing with Expert Systems (ES) in an automated analysis to design a simple and easy-to-use computer interface. Expert systems are a branch of Artificial Intelligence (AI) [3] that aims to serve automated decision-making by sharing knowledge with experts to define a set of rules and steps for making the best decision [4] in an interactive method. This interactive method allows multiple choices in line with the desires of the end users, which makes the decision-maker involved in the decision-making process in an easy and simplified way.

This Interaction is called Goal-Oriented Design which is concerned with meeting the needs and desires of the end users in an easy way and smooth steps, creating Human-Computer Interaction (HCI) and achieving Cognitive Walkthrough among the various variables affecting site selection project [5], and allowing interaction with the experiences and expertise of the end user [6].

E-Marketing, our applied study, is defined as the process of marketing different products or services over the Internet. It is also a broad, integrated concept that includes advertising, communicating with customers/clients, and retaining them for as long as possible. That is, the use of the Internet and other interactive means to build interaction between the company and specific consumers/clients. In other words, e-marketing is also meaning that the company can use the Internet continuously at every step of the sale; before, during, and after the sale. E-marketing is important for companies as it helps them attract the largest possible number of customers *via* the Internet, in addition to advertising and promoting their goods and services online [7]. E-marketing has a wide range of methods, starting with building a company's website on the Internet, passing through electronic advertisements, e-mail marketing, and creating electronic catalogues. Also, E-commerce companies raise their customers' e-loyalty by appropriately managing their electronic word of mouth (EWOM) systems [8]. On the other hand, E-shopping is defined as the process of browsing web pages specialised in selling various goods and services through what is known as electronic shops for electronic purchase by the consumer himself.

1.1 Problem statement

Decision support programmes for the best site selection for any economic investment, E-marketing, for example, are highly specialised programmes that specialists and experts use through a set of specialised steps that the project owners may not be familiar with. Also, decision support programmes may be costly and need a long time to use efficiently. Also, there may be an incompatibility between the results reached by the consultant and the wishes and experiences of the end user/client.

So, through the proposed DSE, the client will be involved in the decision-making process in an interactive way by making him select the important variables affecting his project.

The main contributions of this work can be summarised as follows:

Program a web-based Decision Support Engine (DSE) to facilitate the decision-making process using recent open-source technologies. Some examples of these open-source technologies are Python Programming Language, Django Web Framework, HTML5, CSS, and JavaScript. These open-source technologies are mainly used in building web applications that rely on Python as a backend to apply some data analysis and statistics and display the results on the web in an efficient manner. Our work will use Django Web Framework.

Involve the end user in the decision-making process in an interactive manner.

The rest of this chapter is organised as follows. Section 2 outlines the related work briefly. The proposed methodology is presented in Section 3. Experimental results for the proposed method are introduced in Section 4. Finally, Section 5 concludes the chapter.

2. Related work

The related works could be divided into three groups:-

Group 1 shows the studies of Decision-Making Process. For example, a geospatial Decision Support system (DSS) was implemented to achieve high quality for drought-related data. This DSS offers a range of data mining techniques with full control to users in the United States [9, 10].

Also, a Geospatial Cyber-infrastructure platform was used to build a novel DSS that facilitates collection, management, processing for both static and dynamic datasets, data visualisation, and on-the-fly computer applications in order to perform simulation modelling [11].

The power of automated analysis and Electronic DSS in decision making could be used in medicine to facilitate communication between patients and doctors, in order to reach the best service through applications of information technology [12].

From a computer sciences view and the programming languages, Shesham [13] proposed a framework named GeoFilter which derives GIS data and processes it with Expert System decision-making capabilities to help in solving spatial decision-making problems.

Also, Ahmed et al. [14] show how an information system in Egypt helps investors to take their investment decisions. This system adopted many different technologies, such as JavaScript, Google Maps API, SQL Server, Liferay Content Management System, and JSP. The system was tested by different platforms, such as Microsoft Azure Cloud, VM (Windows 7 based) on Cloud (PaaS Service), and a private remote server. Also, query time at each platform was measured. The results showed that applying the query on a virtual machine on a private cloud was the best, which took the shortest time.

Also, in the field of business and investment, ArcGIS Business Analyst helps to add business, demographic, and consumer data to provide tools for market analysis. This Business Analyst helps in designing campaigns that consider demographics and planned launch dates to help offer insight into designing a strategy for marketing [15].

A retail GIS analysis explained how GIS and some methods and new tools could be used in site suitability analysis, customer profiling, future market potential analysis, and retail sector trends [16].

Other studies show the relationship between Decision-Making Process and planning goals, such as Land Information Systems (LIS), which developed as a tool for urban planning, development, and management on different planning levels. LIS has

been integrated with many GIS platforms for effective implementation [17]. Also, ASPENEX for forest management, URBYS for urban planning and analysis, and GEOOEX to assist planners in assessing the suitability of land for land use activities. These applications are produced by cooperating with GIS and the Computer Science represented in the field of programming.

Group 2 revealed studies that show how Public participation can be used in Decision-Making. For example, the data from geographic GIS and economic modelling can be integrated with a visual MCDM tool to enhance the National Environmental Policy Act (NEPA) process and public participation. Also, user priority is combined with the MCDM tool to provide multiple alignment alternatives for transportation project participants [18]. In addition, public participatory GIS (PPGIS) was introduced to help the public in collaborative spatial decision-making (CSDM) by participating in the site selection process through a web-based platform using cloud computing infrastructures [19].

Group 3 discussed the relationships between the interactive maps by using Web-GIS, and the decision-making process. Web-GIS could be considered an interactive method of programming to achieve the decision-making process. Web-GIS is used in many fields such as tourism, health, economics, management, etc.

For example, ArcGIS Server and Web-GIS are used to create a platform guide for tourists to enable them to find the touristic places and services in historic Cairo in an easy and quick manner. In this study, the digitisation of the cultural heritage led to creating of digital maps for historic Cairo, which led to easy and accurate identification, analysis, and interpretation of geographical data and touristic places attributes [20].

The Zambian Ministry of Health created a web-based GIS architecture with an intuitive and simplified user interface to enable non-specialist users to work on the system without any specialised knowledge. Their architecture was used as a Managing tool for monitoring and evaluating health facilities and as a portal for public interaction with their spatial information [21].

Web-GIS maps were utilised to investigate malaria disease patterns and prevalence related to the environment in the Congo and their relationship with environmental and demographic factors to increase public health literacy through dynamic mapping for timely and accurate decision-making [22].

Also, a web-based GIS system was built for monitoring and mapping the Sugarcane crop at the farm level in Medak District—Andhra Pradesh, India. This monitoring system led to appropriate decisions to increase production and other activities related to the crop. It also helped non-technical users access the information and take appropriate measures to improve crop production [23].

Another Web-GIS was used to build an information system for tourism marketing in the Aqaba Governorate, which resulted in saving time and effort and reaching accurate results in a very short time [24].

On the other hand, a Web-GIS decision support system for rural land management in Wadi El-Natrun Valley (WNDSS), El-Beheria governorate, was designed and implemented. This WNDSS was developed following the client/server model and contained many functions, like data extraction, statistical analysis, and visualisation by an interactive map [25].

Another Web-GIS platform was proposed to monitor and investigate soil at multi-scale in order to raise public awareness. This platform is a combination of Web-GIS with on-the-fly geospatial processing based on GPU computing, specifically designed to allow real-time requests [26].

Three-dimensional interactive maps were introduced to offer improvements and benefits to their traditional counterparts by presenting resulting maps in a 3D view [27].

From the above discussion, it can be noted that Web-GIS is considered as a first step of programming to aid decision-making by using web platforms. And by using GIS, spatial data could be provided to the planners, because GIS has the ability of storing, retrieving, analysing, modelling, and mapping a large volume of spatial data. And by the integration between these maps and its spatial databases, users can create knowledge used in planning and decision-making.

But, the proposed DSE will have a superior feature over the previous works discussed that is it will increase the clients' participation in the decision-making process by programming a simple interface allowing them to select which criteria affect their project and also select the level of this effect in a process called Public Participation GIS (PPGIS). And the resulting map will aid in the decision-making process to identify a suitable place for e-marketing in the study area.

3. Methodology

In this research, the proposed application was built to help the decision-maker in selecting the best places for E-marketing in the study area. The interactive map was applied to a new object in Economic Geography, "E-marketing," and in an important area "Cairo and Giza governorates" which is the capital of Egypt. A detailed description of the proposed methodology and the key technologies used in developing this system are presented in this section. The proposed methodology is shown in **Figure 1**.

The study area of this work is Cairo and Giza governorates because Cairo is the Capital of Egypt, and Giza is Twin City for it, and both of them have 18.7 million persons [28], which represents 19% of Egypt's population, so they form a promising place for E-marketing, or for any other economic projects. A random electronic questionnaire was distributed over the study area to collect the data.

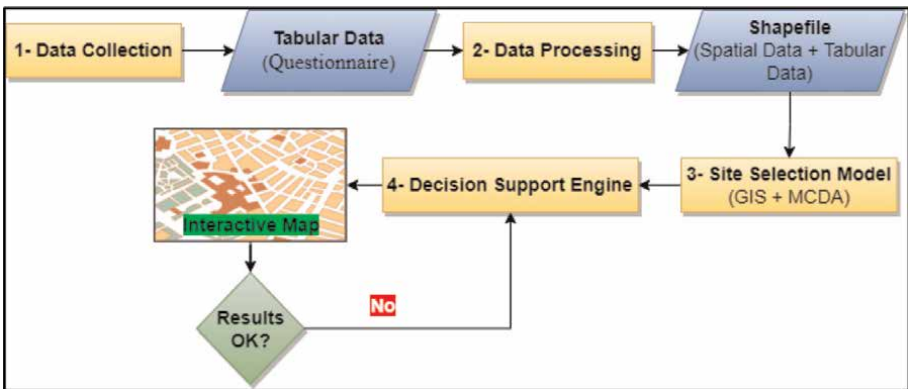


Figure 1.
The proposed methodology.

3.1 Data collection

The first step in the proposed methodology is collecting the data. The data required in this system was gathered by an electronic questionnaire carried out online. This questionnaire (Appendix 1) was designed in Google Forms, and the link to these Google Forms of the questionnaire was sent to the interviewees by Facebook randomly. This questionnaire aims to collect data about the variables that have an effect on e-marketing. These variables were identified and described in general in **Table 1**.

The number of questionnaires collected was 928 questionnaires distributed between Cairo, 363 questionnaires (39.1%), and Giza, 565 questionnaires (60.9%). The outcomes of this questionnaire and the characteristics of the interviewees in classification are shown in **Table 2**, which will be the base for creating geodatabase of the proposed DSE. The classification of every variable into categories is shown in **Table 2**, for example, age group (41–60) represent the high percentage (73.1%) from studying the age variables. Moreover, the university educational level is the high percentage (49.7%) from studying the Education level variables, and so on, see **Figure 2**.

Variable	Description
Age	The younger age most responsive to E-marketing activity
Educational level	The high educational level most suitable for E-marketing
Job	The job types that most interactive with E-marketing
Mother work	The author hypothesis is that The workwoman is more interactive with E-marketing than the unemployed woman according to personal view
Family income	High family income means more activities on E-marketing according to children within the family
Browsing	Internet browsing rate per day
Searching on Internet before shopping	Searching for the desired product online from several websites before buying it

Table 1.
Selected variables of questionnaire.

Variables	Sub-variables	Size	%
Age	≤40	174	18.8
	41–60	678	73.1
	≥61	76	8.2
Education level	Secondary or less	416	44.8
	University	461	49.7
	Above university	51	5.5
Job	Governmental sector	392	42.2
	Private sector	536	57.8
Mother work	Housewife	664	71.6
	Governmental job	174	18.8
	Private sector	90	9.7

Variables	Sub-variables	Size	%
Family Income (EGP)	≤3000	301	32.4
	3001–5000	419	45.2
	≥5001	208	22.4
Daily Internet browsing	≤ 2 hour	263	28.3
	2–4 hours	240	25.9
	≥ 4 hours	425	45.8
Searching on Internet before shopping	Regularly	646	69.6
	Sometimes	282	30.4
Total		928	100.0

Table 2.
Characteristics of the interviewees.

3.2 Data processing

The questionnaire data have been processed in some steps:

- Convert questionnaire data from individual cases (every interviewee represents a case) to spatial cases by using the SPSS programme (cross tabs process).
- Joining these spatial data with the map in ArcGIS to create the shapefile. This shapefile contains spatial attributes, such as the data geometry, as well as the tabular data of the questionnaires. This shapefile is then used as the interactive layer with which the end user will interact. **Figure 3** shows the questionnaire data after processing.

3.3 Site selection model

Site selection model is a common approach to identifying the best place for any project, and assisting with decisions for the best suitable alternative for future site location, factoring in multiple variables of varying importance to identify locations that best meet established criteria for a site. ArcGIS provides a set of tools and functions for users to perform site selection modelling. To identify possible locations, suitability analysis ranks and scores sites based on multiple weighted criteria. First, you will define the problem, identify the criteria for solving it, and generate the input datasets required.

Then, the input values will be transformed to a common ratio or preference scale to compare the criteria. This scale is created according to the author's subjective view, depending on his experiences with the relative importance of every factor and its effectiveness. Next, weights will be assigned to selected criteria and aggregated to determine scores for each potential site. Finally, these scores will be used to rank sites from most suitable to least suitable to identify the best possible locations for the desired application [29].

The site selection model will be built from the cooperation between the academic and experimental views as follows.

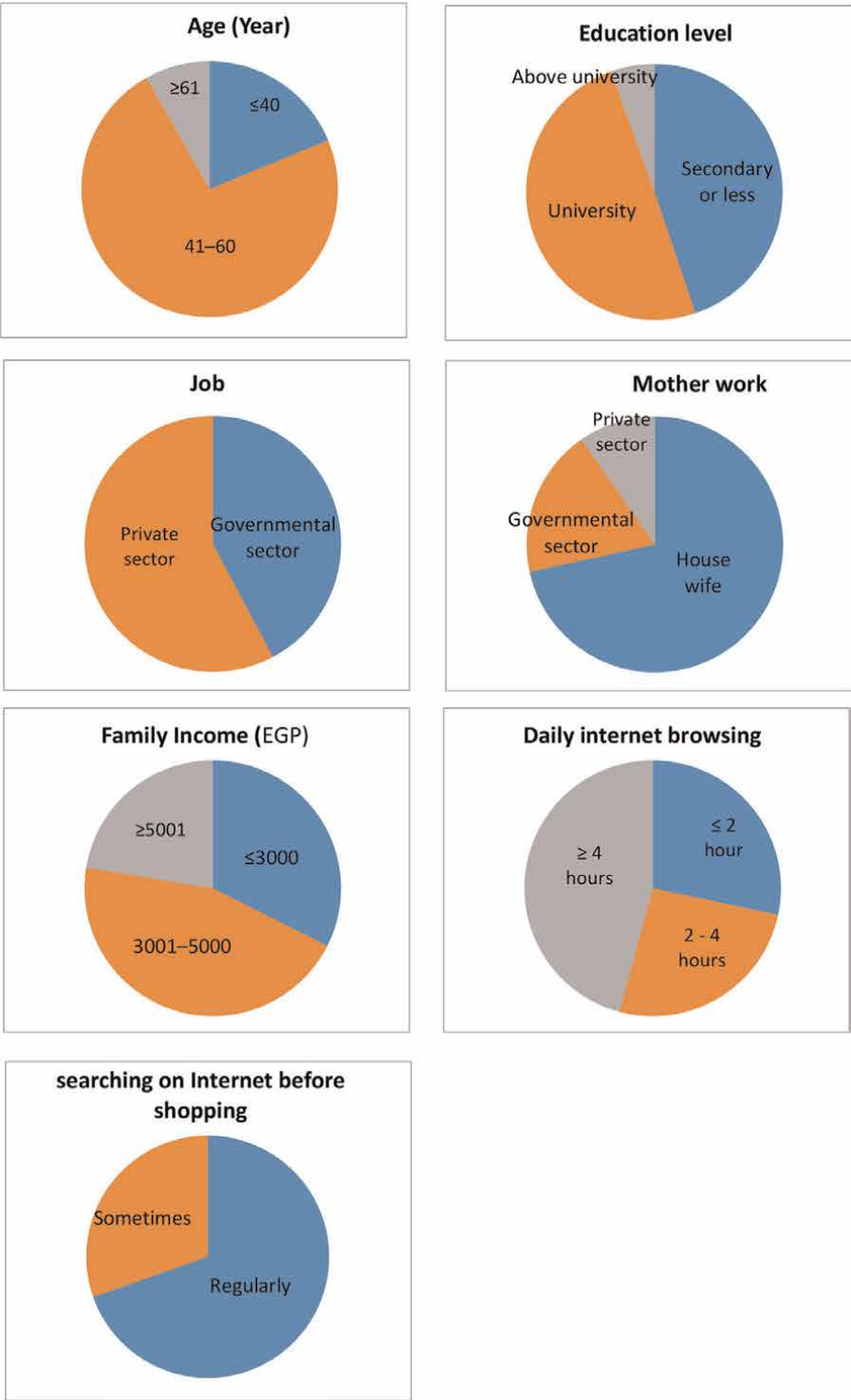


Figure 2.
Characteristics of the interviewees.

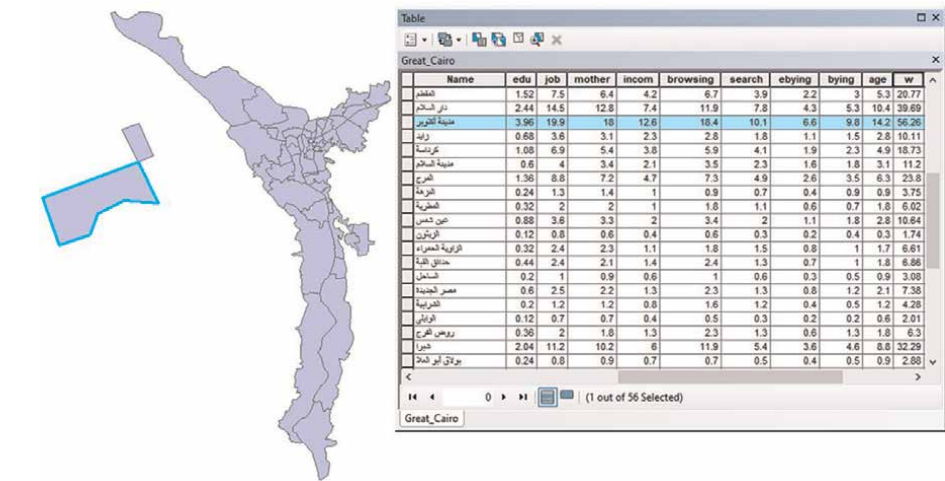


Figure 3.
The questionnaire data in shapefile format.

Variables	Sub-variables	Weights
Age	≤40	0.3
	41–60	0.2
	≥61	0.1
Education level	Secondary or less	0.1
	University	0.2
	Above university	0.3
Job	Governmental sector	0.1
	Private sector	0.2
Mother work	Housewife	0.1
	Governmental job	0.2
	Private sector	0.3
Family Income	≤3000	0.1
	3001–5000	0.2
	≥5001	0.3
Daily Internet browsing	≤ 2 hour	0.1
	2–4 hours	0.2
	≥ 4 hours	0.3
Searching on Internet before shopping	Regularly	0.2
	Sometimes	0.1

Table 3.
Weights of sub-variables in the study area.



Figure 4.
The main interface of the proposed DSE.

- The consultant and the academic will give weights to the sub-variables that have an effect on the e-marketing, depending on his own view, as shown in **Table 3**.
- The end user will put his experiences by selecting the level of effectiveness for every main variable as shown in **Figure 4**. And by this, the clients/end user will participate in the Decision-Making Process of selecting best site of e-marketing in the study area.

3.4 Decision support engine

The last step in the proposed methodology is to build the Decision Support Engine (DSE) by a programming language to identify the best locations for carrying out an E-marketing project. The proposed DSE was built based on free and open-source web framework called Django.

Django is a framework for web development using Python and was created to quickly develop database-driven sites [30]. The Django framework gives a great emphasis on the reuse of models and templates, as well as the DRY (Do not Repeat Yourself) rule [31]. Django is lightweight, efficient, and Python-based. It encapsulates all parts needed to build a web application (Front End and Back End) in one framework, so it simplifies the communications and interaction among the web application components. The front end of the application was built using HTML5, CSS, and JavaScript. It represents the interface with which the end user will interact and is responsible for sending the end user's request to the back end of the process. The back end represents the main controller of the web application, which will receive the end user's requests from the front end, process them, and deliver the results to the front end to render. The back end of the proposed DSE makes use of Python programming language.

4. Results and discussion

The main interface of the proposed DSE is shown in **Figure 4**. In **Figure 4**, all variables that have an impact on the E-marketing activity appear as well as the degrees of their impact to identify the best place for E-marketing in our study area. The proposed DSE provides an interactive process between the decision maker (end user/

clients) and the available variables, where the end user will do the following, according to his own opinion, due to his experiences:

- Determining the variables that have an effect on the project, and leaving the others that do not. These variables are shown in **Table 1**.
- Determining the degree of the impact of each variable on e-marketing (strong—medium—weak) because all variables do not have the same degree of impact. Each factor differs in the degree of its impact on E-marketing according to the opinion of every end user.

After selecting the desired variables as well as the degrees of their impact—as a random example—as shown in **Figure 5**, a site selection model equation will be calculated within the proposed DSE to find the relationship between the selected variables according to the degree of effect of each variable separately, which have subjective reference scale as follows:

If the degree of effect is chosen “strong,” the result of the variable will be multiplied by an effect factor of 1.

If the degree of effect is chosen as “medium,” the result of the variable will be multiplied by the effect factor of 0.6.

If the degree of effect is chosen as “weak,” the result of the variable will be multiplied by the effect factor of 0.3.

All these selections are then translated in the proposed DSE to calculate the site selection model equation, and Algorithm 1 shows how these variables are translated into the site selection model equation.

Taking into consideration the probability of changing these arithmetic transactions from one place to another, and from time to time, the final result of the proposed DSE is an interactive map which shows the selection of some variables (randomly) and the degree of strength of their impact on e-marketing. It also shows the best e-marketing places in Cairo and Giza. The best locations are shown and ranked in categories based on the suitability colour ramp. **Figures 5** and **6** show two different scenarios of the resulting interactive map. The results of every scenario differ according to the variables that were selected by the end user. Hence, the end user is involved in the

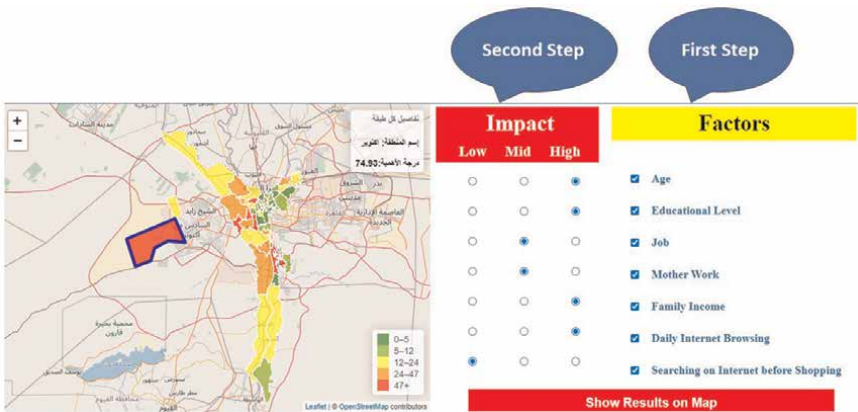


Figure 5.
First scenario of interactive map of E-marketing in Cairo and Giza governorates.

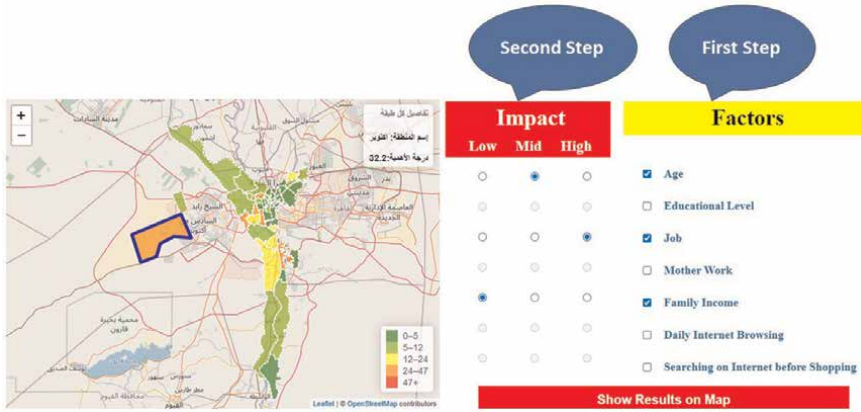


Figure 6.
Second scenario of interactive map of E-marketing in Cairo and Giza governorates.

decision-making process because he can select the most important variables and not select the other, according to his own opinion.

So, the proposed DSE can be thought an integration between the end user's experience and the academic methodology in determining the best locations for E-marketing because the site selection model will be built from the cooperation between the academic and experimental views. The consultant and the academic view will give weight to every sub-variables that affect the e-marketing, depending on the author's view. After that, the end user will select the level of effectiveness for every main variable. And by this, the clients will participate in the Decision-Making Process of selecting the best site of e-marketing in the study area.

This integration must be done because the site selection process varies between the specialist/consultant applying the academic experience and the client/end user with practical experience. Hence, the proposed DSE will allow the end user to participate or determine the variables that affect e-marketing through cooperation with the academic specialist.

Algorithm 1. Site selection model calculation.

```

if request.is_ajax and request.method == "POST":
    try:
        age = float(request.POST.get('age'))
    except:
        age = 0.0
    try:
        edu = float(request.POST.get('edu'))
    except:
        edu = 0.0
    try:
        job = float(request.POST.get('job'))
    except:
        job = 0
    try:
        mother = float(request.POST.get('mother'))
    except:
        mother = 0.0

```

```
try:
    income = float(request.POST.get('income'))
except:
    income = 0.0
try:
    internet = float(request.POST.get('internet'))
except:
    internet = 0.0
try:
    buy = float(request.POST.get('buy'))
except:
    buy = 0.0
total_weight = (feature.GetField("age") * age) + (feature.GetField("edu") * edu) +
(feature.GetField("job") * job) + (feature.GetField("incom") * income) +
(feature.GetField("browsing") * internet) + (feature.GetField("mother") * mother) +
(feature.GetField("search") * buy)
```

As a result of this integration, the end user will participate in the decision-making process in a no-cost, easy, and fast manner, which will increase the utilisation of new technologies that facilitate communication between academics and end users to save time and effort of discussion between them. And, the user experience (UX) is not necessary, because the interface of the application is very simple and easy to use. So, these processes will make the communication between the academic and practical views more active and dynamic. Thus, the proposed DSE will be a modern technological tool that the economic community can benefit from, to develop their projects and facilitate the communication process between the investors and the academics, and to exchange experiences with each other, thus supporting the final decision-making.

As time means money, a practical benefit of the proposed DSE is the fast, economical proper decision-making to find the best location for e-marketing and other economic projects, which means increasing profit compared to competitors.

And over time, the proposed DSE can be a commercial application by being an economic advisory tool available online. Also, it may have advertising revenues if it has many visitors.

The proposed DSE also can be considered as a base for other electronic applications and for more developed and detailed ideas, for various economic projects, through the following:

Using more advanced tools like advanced mathematical and statistical methods, such as relative weights methods and others, leads to more accurate and comprehensive Decision-making Modelling [32] for e-marketing projects or similar projects.

Providing the application with more detailed data (big data) by using more advanced technologies [33] about the population, lifestyle and habits of shoppers and customers for the e-marketing projects or other similar projects, to benefit the investors in determining the best sites of marketing a new product.

Update the application's data with a time series forms, to make a comprehensive view.

5. Conclusion

Decision support programmes for the site selection process in E-marketing are specialised programmes that specialists and consultants use. In this chapter, a Decision

Support Engine was proposed by a programming language to facilitate the decision-making process and involve the end user in such an interactive manner.

An online electronic questionnaire gathered the data to create ArcGIS geodatabase for our platform; Django web framework was used for programming and building this Decision Support Engine. The end result was an Interactive Map to Support Clients' Decision-Making of E-marketing in Cairo and Giza Governorates in an easy manner. In the future, variables inside our application can be updated and developed to make the resulting decision and map more accurate and comprehensive. It can also predict its future using prediction systems and geographical attitudes, forecasting its potential models and scenarios.

Using the proposed DSE, the client will be involved in the decision-making process through the selecting variables that have an impact on the site selection process for E-marketing in Cairo and Giza governorates by an interactive method to produce an interactive map showing the best location of e-marketing through direct participation of the clients/end users.

In future, the raw data of the DSE can be updated and developed to make the decisions made more comprehensive. Also, it can be extended by using prediction systems, which allows the DSE to forecast future scenarios from a geographical point of view.

Acknowledgements

The authors were very grateful to the student of Geography Department, Faculty of Arts, Cairo University about their helping in achieving the questionnaires.

Appendix

Questionnaire for the research of “The interactive Map of E-marketing to Enhance Clients' Participation in Decision-Making Process, in Cairo and Giza Governorates-Egypt.”

All replies will be secret

Age: (... . Years).

Education level: Secondary or less() University() Above university().

Job: Governmental sector() Private sector() Governmental in the morning and private afternoon() Handicraftsman().

Mother work: House wife() Governmental job() Private sector().

Family Income: Less than 3000 EGP () 3000–5000() More than 5000 EGP ().

Daily browsing: Less than 2 hours() 2–4 hours() More than 4 hours().

Internet shopping: Regularly() Sometimes().

Author details

Reda Alkot Mohamed^{1*} and Sayed Ahmed Sayed^{2,3}

1 Faculty of Arts, Geography Department, Cairo University, Giza, Egypt

2 Faculty of Graduate Studies for Statistical Research, Department of Computer Sciences, Cairo University, Giza, Egypt

3 National Authority for Remote Sensing and Space Sciences, Egypt

*Address all correspondence to: redaelkot@cu.edu.eg

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Eldrandaly K, AbdelAziz N. Enhancing ArcGIS decision making capabilities using an intelligent multicriteria decision analysis toolbox. *Journal of Environmental Informatics*. 2012;**20**:44-57
- [2] Chandio I, Matori A. GIS-based multi-criteria decision analysis of land suitability for hillside development. *International Journal of Environmental Science and Development*. 2011;**2**: 469-473
- [3] Drake R, Deegan P, Woltmann E, Haslett W, Drake T, Rapp C. Comprehensive electronic decision support systems. *Psychiatric services*, ps.psychiatryonline.org. 2010; **61**:714-717
- [4] Robinson V, Frank A. Expert Systems for Geographic Information Systems. *photogrammetric Engineering and Remote Sensing*. 1987;**53**:1435-1441
- [5] Roth R. Interacting with Maps: The Science and Practice of Cartographic Interaction [Thesis]. The Graduate School College of Earth and Mineral Sciences. USA: Department of Geography: The Pennsylvania State University; 2011
- [6] Lee S, Shim J. Interactive goal programming on the microcomputer to establish priorities for small business. *Journal of the Operational Research Society*. 1986;**37**:571-577
- [7] Kolomvatsos K, Anagnostopoulos C, Hadjiefthymiades S. Sellers in e-marketplaces: A fuzzy logic based decision support system. *Information Sciences*. 2014;**278**:267-284
- [8] Khan M, Hashmi H. Impact of interactivity of electronic word of mouth systems and website quality on customer E-loyalty. *Pakistan Journal of Commerce and Social Sciences*. 2016;**10**:486-504
- [9] Goddard S, Harms S, Reichenbach S, Tadesse T, Waltman W. Geospatial decision support for drought risk management. *Communications of the ACM*. 2003;**46**:35-37
- [10] Harms S. Data mining in a geospatial decision support system for drought risk management. In: *Proceedings of the 1st National Conference on Digital Government*. California, USA: Digital Government Society of North America; 2002. pp. 9-16
- [11] Terribile F, Bonfante A, D'Antonio A, De Mascellis R, De Michele C, Langella G, et al. A geospatial decision support system for supporting quality viticulture at the landscape scale. *Computers and electronics in agriculture*. 2017;**140**:88-102
- [12] Beilby J, Duszynski A, Wilson A, Turnbull D. Electronic decision support systems at point of care: Trusting the deus ex machine. *MJA*. 2005;**183**:99-100
- [13] Shesham S. Integrating Expert System and Geographic Information System for Spatial Decision Making [Thesis]. USA: Department of Computer Science, Western Kentucky University; 2012
- [14] Ahmed S, Badawy M, Zidan E, Ibrahim R. Implementation of an investment information system based on Google maps API. *International Journal of Advanced Research in Computer and Communication Engineering*. 2015;**4**: 383-388
- [15] Connor J, Watson K, Anders A, Berry M, Tsushima P, Kaplan D. Using

ArcGIS Business Analyst. USA: ESRI; 2008

[16] Kavita K, Patil G. Geographic information system (GIS) – For business analytics. *International Journal of Scientific & Engineering Research*. 2011; 2:1-6

[17] Sukeerthi N. Land Information System as a Tool for Urban Planning: Integrated Land Information System in GIS Platform. LAP LAMBERT Academic Publishing; Illustrated edition; 2013

[18] Bethany S, Joseph H. Using a multiple criteria decision-making model to streamline and enhance NEPA and Public participation processes. *Public Works Management & Policy*. 2011;16: 59-89

[19] Farnaghi M, Mansourian A. Blockchain, an enabling technology for transparent and accountable decentralized public participatory GIS. *Cities*. 2020;105:102850

[20] Mohammed R, ABD EL Gawad Z, Voda M. Using web GIS for marketing historical destination Cairo, Egypt. *Geographia Technica*. 2021;16:193-204

[21] Mushonga H, Banda F, Mulolwa A. Development of a web based GIS for health facilities mapping, monitoring and reporting: A case study of the Zambian ministry of health. *South African Journal of Geomatics*. 2017;6: 321-332

[22] Reed R, Bodzin A. Using web GIS for public health education. *International Journal of Environmental and Science Education*. 2016;11:6314-6333

[23] Kumar S, Babu S. A web GIS based decision support system for agriculture crop monitoring system-A case study

from part of Medak District. *Journal of Remote Sensing & GIS*. 2016;5:1-19

[24] Ahmed A. Building a Marketing Information System for Tourism in Al Aqaba Using ARCGIS 9.3 Software [Thesis]. Jordan: Business Management Department: Middle East University; 2011

[25] Ahmed S, Kotb A, Samir E, Moustafa M, Farg E, Abd Elhay A, et al. A WebGIS decision support system for Wadi El Natrun rural land management. In: *Proceedings of 21st International Arab Conference on Information Technology (ACIT)*. Egypt: IEEE; 2020. pp. 1-7

[26] Langella G, Basile A, Giannecchini S, Moccia F, Mileti F, Munafó M, et al. Soil monitor: An internet platform to challenge soil sealing in Italy. *Land Degradation & Development*. 2020;31: 2883-2900

[27] Góralski R. Three-Dimensional Interactive Maps–Theory and Practice [Thesis]. United Kingdom: University of South Wales; 2009

[28] CAPMAS. Egypt in Figures, Population Estimation in January 2019 [Internet]. 2021. Available from: <http://www.capmas.gov.eg/Pages/populationClock.aspx> [Accessed: October 1, 2021]

[29] Abramovich A. Using GIS to Assist Location and Site Selection Decisions [Thesis]. Urban Design and Planning Program. USA: University of Washington; 2012

[30] Liawatimena S, Warnars S, Trisetyarso A, Abdurahman E, Soewito B, Wibowo A, et al. Django web framework software metrics measurement using radon and pylint. In: *Proceedings of Indonesian Association*

for Pattern Recognition International Conference (INAPR). USA: IEEE; 2018. pp. 218-222

[31] Aziz E. Design architecture of an integrated student activities management system for higher education. Turkish Journal of Computer and Mathematics Education (TURCOMAT). 2021;**12**:1676-1683

[32] Abdul RM. Capabilities of a GIS-based multi-criteria decision analysis approach in modeling migration. GeoJournal. 2019;**84**:483-496

[33] Rowe F. Big data and human geography. In: Demeritt D, Lees L, editors. Concise Encyclopedia of Human Geography. Edward Elgar Encyclopedias in the Social Sciences Series. 2023

Visualization and Machine Learning Prediction of Spatiotemporal Spread of COVID-19 in India

Rifaat Abdalla and Imen Hamdi Nasr

Abstract

The global COVID-19 pandemic, affecting over 8 million people across 100 nations, presents a severe risk to human life and property. India, with its vast population of 1.34 billion, is among the hardest-hit countries. This study employs machine learning techniques for visualizing and predicting the spatiotemporal progression of COVID-19. Utilizing Python libraries such as “Numpy,” “Pandas,” “Scikit,” and “Matplotlib,” we analyze and visualize COVID-19 data sourced from the Indian Ministry of Health Web Service and API. Our visualizations depict demographic trends, incident growth, geospatial state-based patterns, and distribution. The analysis reveals that age groups under 30 and over 59 exhibit resilience to the virus, offering hope for population growth. Examining active cases, recoveries, and deaths, India has outpaced countries like Germany, the United States, Iran, Italy, Spain, South Korea, Turkey, France, and the United Kingdom since early April 2020. Furthermore, we employ supervised machine learning algorithms, including PROPHET and ARIMA, to predict the virus’s spread. By accounting for seasonality-related factors, we achieve a 95% prediction interval, indicating the potential for accurate spread forecasting. This research contributes valuable insights into COVID-19’s impact in India and offers predictive tools for managing its progression.

Keywords: pandemic, COVID-19, geospatial, data modeling, data visualization, machine learning, PROPHET, ARIMA, India

1. Introduction

The coronavirus (COVID-19) pandemic is the biggest humanitarian crisis to hit the planet since World War II. The pandemic virus has widely spread, and there are more cases being reported every day [1]. The federal government is attempting to stop its spread. Till date it has spread across 215 countries infecting 5,491,194 people and killing 346,331 till May 23, 2020. As of May 23, 2020, there have been 138,536 COVID-19 cases registered in India. Of these, 57,692 patients have recovered and 4024 have died.

On January 30, 2020, the state of Kerala in India reported the country's first confirmed case of the coronavirus illness [2]. The impacted person had traveled previously from Wuhan, China. More than 2000 coronavirus cases have been confirmed in India as a result of testing thousands of suspected cases. Maharashtra, Kerala, Delhi, Karnataka, Andhra Pradesh, Uttar Pradesh, Rajasthan, and Tamil Nadu are the states with the most cases. The first coronavirus-related mortality in India was noted in Karnataka on March 12. By the end of March 2020, 52 deaths had been recorded nationwide.

India took swift action by enacting a proactive, statewide shutdown to flatten the curve and give itself time to properly prepare and fund its responses. India has reported 3720 deaths from 32 States and Union Territories as of May 23rd, 2020, and there have been a total of 123,202 confirmed cases of COVID-19 [3]. In order to assess, forecast, and implement safety measures against COVID-19 in India, data scientists throughout the world are employing AI and machine learning. (Reference). India is preparing for the COVID-19 outbreak in accordance with official government directives, but failing to take certain crisis measures or underestimating their significance may have very serious repercussions. India's surrounding nations have all reported COVID-19 positive cases. **Figure 1** depicts the timing of COVID-19's propagation in India from January 30 to March 25, 2020.

The aim of this research is to explore the COVID-19 situation in India and the world in the coming 15 days, through a machine learning model that projects how the virus might spread throughout India. Specifically, this paper will analyze the condition in India from January 30, 2020, to May 24, 2020, and provide a prediction model for the following 15 days. In doing so, the research scraps out COVID data from the Indian Ministry of Health website, through the open API, and provides visualization for the mortality rate per 100 across the affected states. This paper provides visualization for demographics, age groups of affected patients, and a number of tests performed along with state-wise testing insight and visualization of ICMR testing centers in each state. Finally, Auto Regressive Integrated Moving Average "ARIMA" prediction was performed to forecast confirmed cases in India. The final stage was to compare the Indian cases of COVID-19 along with the global spread of the pandemic and provide a geospatial visual model for the hotspots and affected regions of India. The study will also compare the data obtained about the expansion of COVID-19 in India and globally. Numerous researchers, including [4, 5], have looked into the spatial distribution of COVID-19.



Figure 1.
The early explosion of COVID-19 in India.

2. Literature review

The study of giving machines and computers the ability to learn without being explicitly programmed through the exploration of algorithms that can self-teach and make predictions about data falls under the umbrella of machine learning, a subfield of computer science that emerged from the field of artificial intelligence. Instead of executing machine-bound, static, and code-defined instructions, these algorithms work by creating models of inputs to aid in decision-making [6].

Machine learning is defined as a set of methods that can detect hidden patterns in data automatically without explicitly programming them and then use these hidden models to predict future data or assist in making decisions. Machine learning is closely related to the fields of statistics, pattern recognition and data mining, a field of research that enables computers to learn without explicit programming [7, 8].

One method of teaching software to learn and improve automatically through experimentation—even without being expressly coded or changed by its programmers—is through machine learning. Examples of training data, or data used in the creation of mathematical models by algorithms.

Machine learning depends on quickly analyzing a vast amount of data, and then can be linked to the decision-making processes and predict the future, where the computer analyzes a huge amount of human failure in normal ways to analyze and by studying the patterns and trends of data Known as Predictive Analysis, apart from the precise technical aspects of machine learning or learning derived from information systems, the uses and applications of these technologies are enormous and extremely useful in various fields, and contribute significantly to decision-making processes and save effort and time with Mechanism of accuracy [9].

Supervised Learning, Unsupervised Learning, Semi-Supervised Learning, and Reinforcement Learning are the several types of machine learning algorithms (Figure 2).

Its applications are used in agriculture, irrigation, and in the fourth generation of industry, communications, roads, mines and health care, and are used in the fields

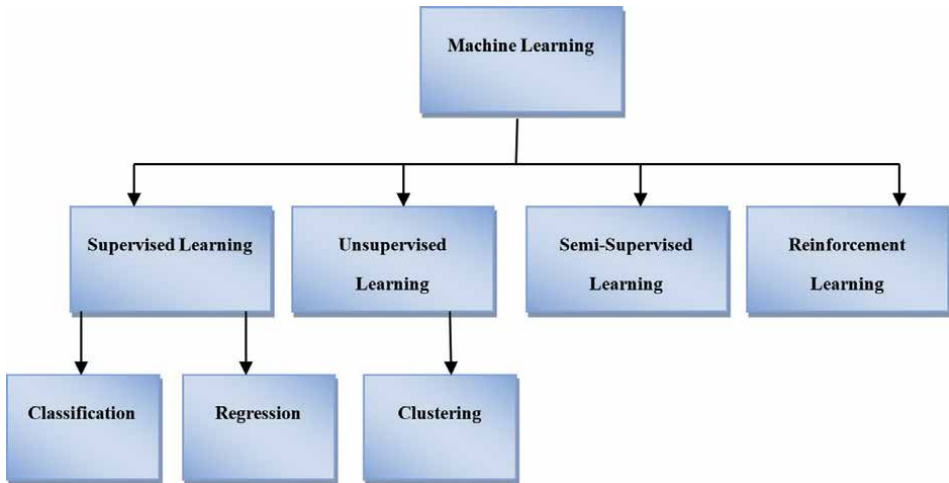


Figure 2.
Classification of machine learning approaches.

of marketing, insurance, banking and translation [10]. For example, in agriculture, these techniques are used in crop management in so-called precision agriculture. Plants or trees individually, through simple sensors can know the tree is treated and the exact dose required, as well as irrigation in time with the amount of water that fits perfectly, which contributes directly to increasing production and saving a lot of time and Ward even predicts the disease before it occurs, in addition to conducting all tests and examinations without human intervention and take appropriate decisions in a timely manner [11]. Machine learning has been utilized extensively most recently during the COVID-19 spread [12–14].

3. Methodology

The flow process of the design of machine learning protocol, in general, starts initially with understanding the problem and determining the goal of the design of the model or algorithm. The second stage is to collect and categorize the data if not collected before. Followed by cleaning the data if there are problems. The important stage is the model design, followed by training the model after selecting the appropriate model, and evaluate the results of the model, and finally put the model for use by the user (**Figure 3**).

The application of Machine Learning for visualization and prediction is growing. Through Python Programming Language and a number of dedicated libraries, it is possible to achieve advanced visualization including the Python “Pyplot” Toolkit, which is a set of functions that make “matplotlib” function like MATLAB. Matplotlib is a comprehensive library for building static, animated, and interactive visualizations. Each “pyplot” function modifies a figure in some way, such as by creating a figure, a plotting region within the figure, some lines within the plot “Seaborn” is aarea, and labeling the plot. A Python data visualization library called “Seaborn” is based on” “matplotlib.” The Plotly “Express” high-level API for making figures is clear, consistent, and provides a high-level interface for creating visually appealing and educational statistical “Express” is a concise, consistent and high-level API for creating figures. Visuals. The graph objects in “Plotly.graph” objects” are the images

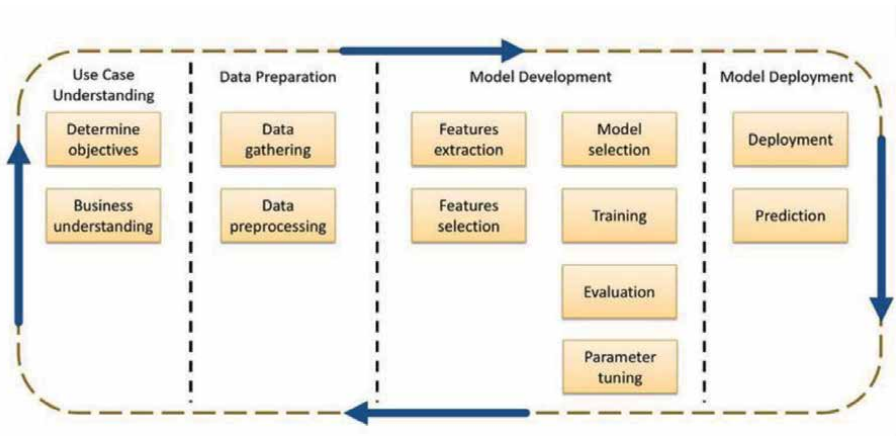


Figure 3.
Stages of work in the design of machine learning models.

```
# importing the required libraries
import pandas as pd

# Visualisation libraries
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
import folium
from folium import plugins
```

Figure 4.
Visualization libraries that are useable in python.

produced by the Plotly Python package. Folium library provides mapping functions that is builds on the data wrangling strengths of the Python and allow for mapping capabilities (**Figure 4**).

A class of models known as Auto Regressive Integrated Moving Average (ARIMA) deals with time series using the model's historical values. The ARIMA model is ranked according to the Auto Regression (p), Moving Average MA, and the quantity of differencing necessary to make the time series stationary. In order to estimate future values, it relies on model lags and lagged forecasting mistakes. All-time series that show patterns, are not random, and do not have a lot of noise are suitable for ARIMA 'models. 'explains' a given time series using its own lags and prediction errors, or its own past values, so that the equation can be used to predict future values (**Figure 5**).

The Core Data Science team at Facebook has published Prophet as open source software.

```
from statsmodels.tsa.arima_model import ARIMA

from datetime import timedelta

arima = ARIMA(data['y'], order=(5, 1, 0))
arima = arima.fit(trend='c', full_output=True, disp=True)
forecast = arima.forecast(steps= 30)
pred = list(forecast[0])

start_date = data['ds'].max()
prediction_dates = []
for i in range(30):
    date = start_date + timedelta(days=1)
    prediction_dates.append(date)
    start_date = date

plt.figure(figsize= (15,10))
plt.xlabel("Dates",fontsize = 20)
plt.ylabel('Total cases',fontsize = 20)
plt.title("Predicted Values for the next 15 Days" , fontsize = 20)

plt.plot_date(y= pred,x= prediction_dates,linestyle ='dashed',color = '#ff9999',label = 'Predicted');
plt.plot_date(y=data['y'],x=data['ds'],linestyle = '-',color = 'blue',label = 'Actual');
plt.legend();
```

Figure 5.
Code line showing the of implementing ARIMA model.

Downloadable versions are offered. When forecasting time series data using an additive model, Prophet was utilized to fit non-linear trends including weekly, daily, and yearly seasonality as well as holiday impacts. Strongly seasonal time series and multiple seasons of historical data are ideal for it. Prophet typically manages outliers well and is robust to missing data and changes in the trend. Using Prophet, forecasting confirmed cases of NCOVID-19 for a week in advance with a 95% prediction interval requires building a base model without adjusting any parameters linked to seasonality or adding any new repressors. It offers a precise and quick method for creating trustworthy forecasts for planning and goal-setting. Additionally, it is totally autonomous and highly reliable, giving an accurate forecast on complicated data without any manual work. Prophet can withstand outliers, missing data, and significant time series changes. It offers tunable forecasts for Python and gives customers a lot of options for modifying and adjusting projections.

4. Results and discussion

This section provides an overview of the steps taken in this research. The first part shows the visualization of demographics distribution in India and how the spread of COVID-19 provides some trends and indicators of the impact. The second part shows the preparedness of healthcare facilities and the capacities of urban and rural centers. The third part provides an insight into the testing distribution and pandemic patterns that were in place during the study period. The fourth part provides an insight on modeling and prediction of geospatial trends as it relates to India. The fifth part is showing global modes of the spread of the COVID-19.

4.1 Visualization of demographics distribution

‘Matplotlib’ library was used with Python to visualize the spread of Covid-19 in India. It allowed to visualize the frequency distribution of the numeric array as shown in **Table 1** which is used to show the states, the number of deaths, the number of treated patients, and the number of confirmed cases. The data collected allowed researchers to determine that Maharashtra state had the greatest death rate, confirmed cases, and healed patients, however it was ranked the sixth following the states of lower records in Kerala and Rajasthan, which have provided 4 deaths in total in the state of Karla and 160 cases in the state of Rajasthan, bringing this state to the second position of the reported cases. **Figure 6** is showing the trend in the confirmed cases, cured cases and deaths as a result of Covid-19 in India. **Figure 7** provides further demographic distribution of the cases reported based on the demographics, where it was found that the majority of the reported cases are for the age group between 20 and 29 years old, which represents the highest fraction of the total population, with 24.9% of the total reported sick cases. The second largest group of reported cases is for the age group between 30 and 39 years, with a total of 21% from all reported cases. It was observed that elderly reported cases was in the fifth category for the age group between 60 and 69 with total of 12.9% only surpassing the age group of 50–59 with 11.9% of the total population. The demographics distribution figures confirms that the age group <40 is the most affected which is against the trend which says elderly people are more at risk of being affected. Only 17% of people >60 are affected.

State/Union territory	Confirmed	Cured	Deaths
0. Kerala	795	515	4
1. Rajasthan	6742	3786	160
2. Madhya Pradesh	6371	3267	281
3. Uttar Pradesh	6017	3406	155
4. Goa	55	16	0
5. Ladakh	49	43	0
6. Maharashtra	47,190	13,404	1577
7. Jharkhand	350	141	4
8. West Bengal	3459	1281	269
9. Andaman and Nicobar Islands	33	33	0
10. Assam	329	55	4
11. Manipur	29	4	0
12. Andhra Pradesh	2757	1809	56
13. Puducherry	26	10	0
14. Uttarakhand	244	56	2
15. Bihar	2380	653	11
16. Chandigarh	225	179	3
17. Chhattisgarh	214	64	0
18. Punjab	2045	1870	39
19. Dadar Nagar Haveli	2	0	0
20. Karnataka	1959	608	42
21. Tripura	189	153	0
22. Himachal Pradesh	185	61	3
23. Telangana	1813	1065	49
24. Jammu and Kashmir	1569	774	21
25. Tamil Nadu	15,512	7491	103
26. Meghalaya	14	12	1
27. Gujarat	13,664	6169	829
28. Delhi	12,910	6267	231
29. Odisha	1269	497	7
30. Haryana	1131	750	16
31. Sikkim	1	0	0
32. Mizoram	1	1	0
33. Arunachal Pradesh	1	1	0

Table 1.
 Showing the distribution of confrimed and cured cases and deaths in Indian states.

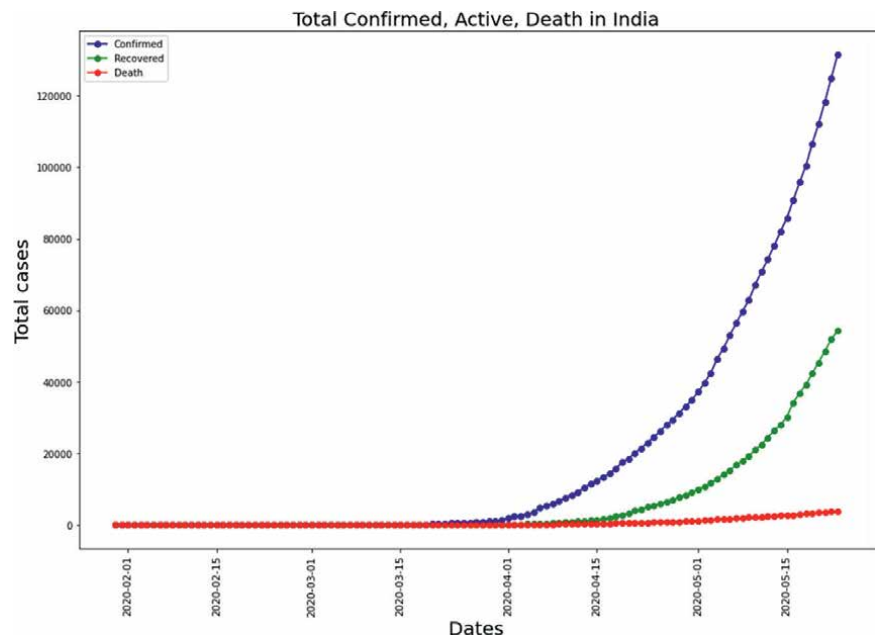


Figure 6.
Visualization of the impact of the COVID - 19 in Indian population as of may 24, 2019.

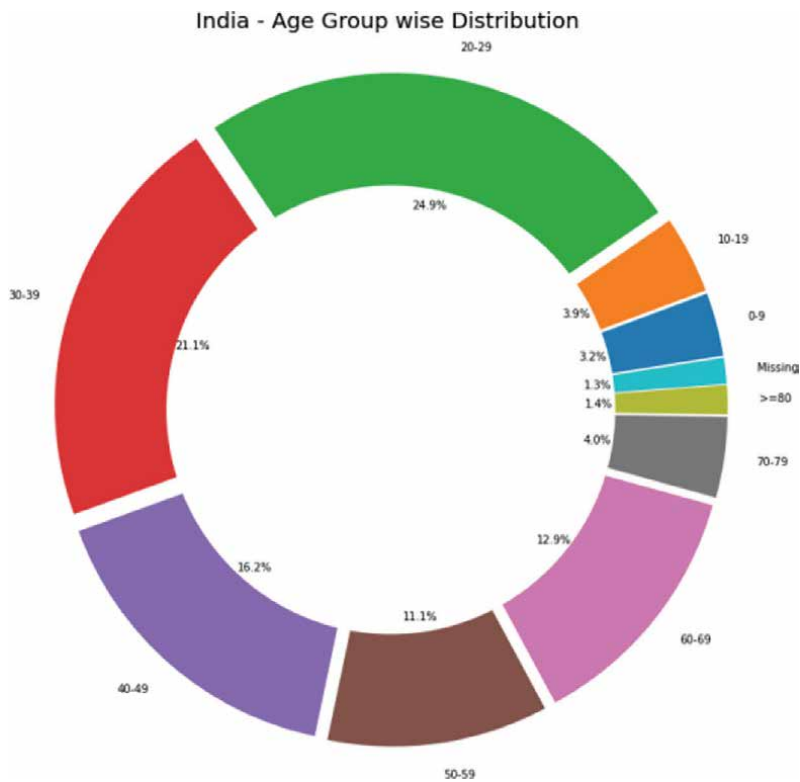


Figure 7.
Age-based demographic distribution of India's population.

4.2 Visualization of the healthcare facilities preparedness

Through the visualization of the rural health care facilities, it was possible to visualize the top ruler hospitals where Tamil Nadu State came with the highest number of urban hospitals beds with beds over 500. In the rural hospitals, the State of Uttar Pradesh came first with a number of hospital beds that exceeds 4500 beds. **Figure 8** provides details of the top 10 urban and rural healthcare facilities distribution and the number of available hospital beds in each state.

A trend was depicted in the distribution of the primary health centers, as the top 10 primary health centers and the top 10 community health centers, as well as the top 10 rural hospitals occur in the Uttar Pradesh State. This may be linked to good infrastructure of this state. **Figure 9** provides an overview of the distribution of health care facilities, in India and how the number of hospital beds distribution, as well as the distribution of facilities in the country. This provides direct association to the useability of these facilities and capacities in dealing with the spread of Covid-19.

4.3 Visualization of COVID-19 testing and pandemic patterns

Testing has been performed at a very limited scale in India. It has been performed through the Ministry of Health, and through the Indian Council of Medical Research (ICMR). Maharashtra state was the most tested by the ICMR, while Lakhra was the least tested. Ministry of health testing was performed starting with one hundred thousand tests in mid-March to a half million test in late April as shown in **Figure 10**. The sampling protocol for ICMR was mainly research oriented to collect samples for

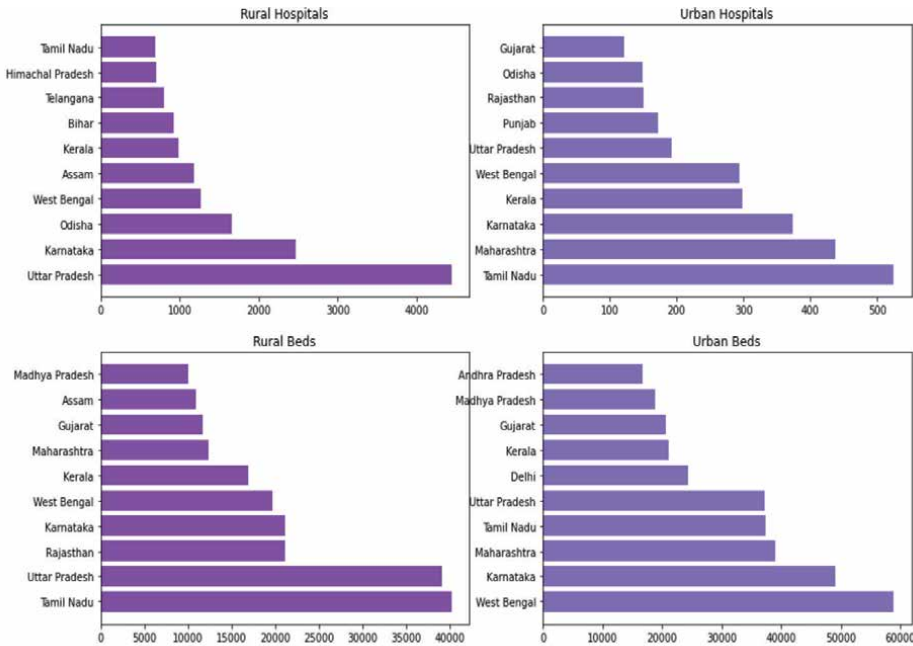


Figure 8.
Capacity of health care facilities and number of beds in urban and ruler India.

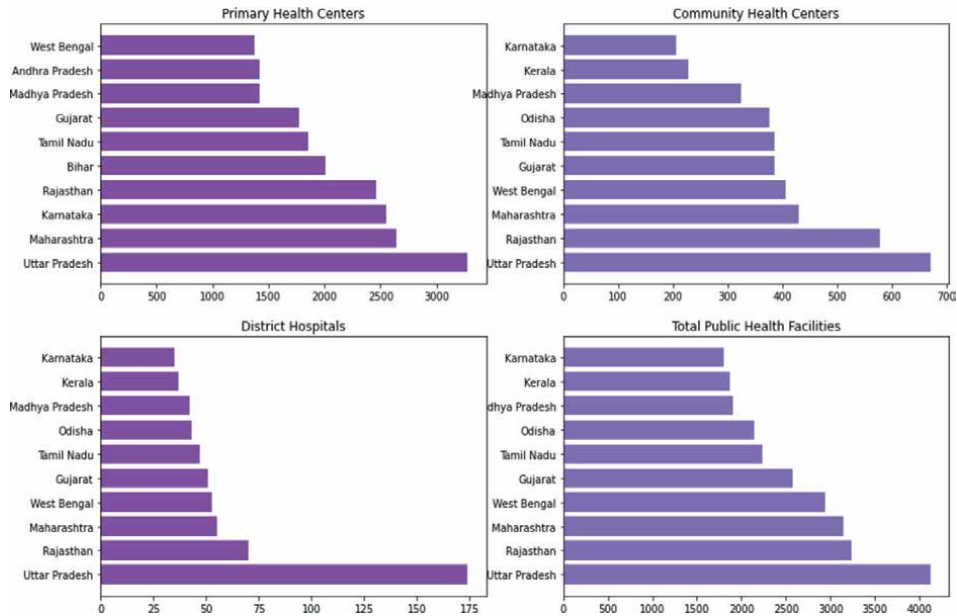


Figure 9.
Distribution of primary and community Health-care facilities in top Indian states.

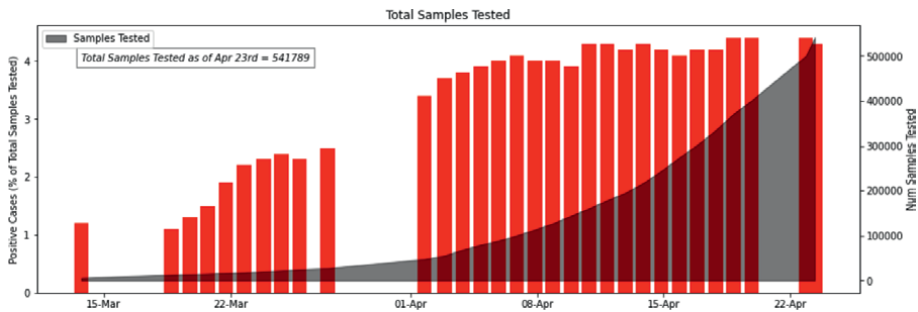


Figure 10.
CPR sample tests in India between march 10 and April 25.

research purpose, the total number of samples collected are shown in **Figure 11**. The distribution of total tested samples, along with negative and positive results is shown in **Figure 12**.

Although being highly populated the relative confirmed cases of India is low compared to other countries. This could be because of two reasons: (1) 67 days lockdown imposed by prime minister Narendra Modi in several stages (Source: Health Ministry) and (2) Low testing rate (Source: news18). Still, the response of the Indian authorities to the pandemic was a matter of appreciation of the WHO and the international community. **Figure 13** is showing the track of confirmed cases weekly in the topmost affected states. The overall pattern is showing that there was a noticeable increase in day-by-day reported confirmed cases between March and May. **Figure 14** is showing the growth rate pattern in the top most affected Indian states for the same period. Similar to spatial analysis of the study area shown by [15], **Figure 15** is showing geospatial visualization of spread hotspots in India.

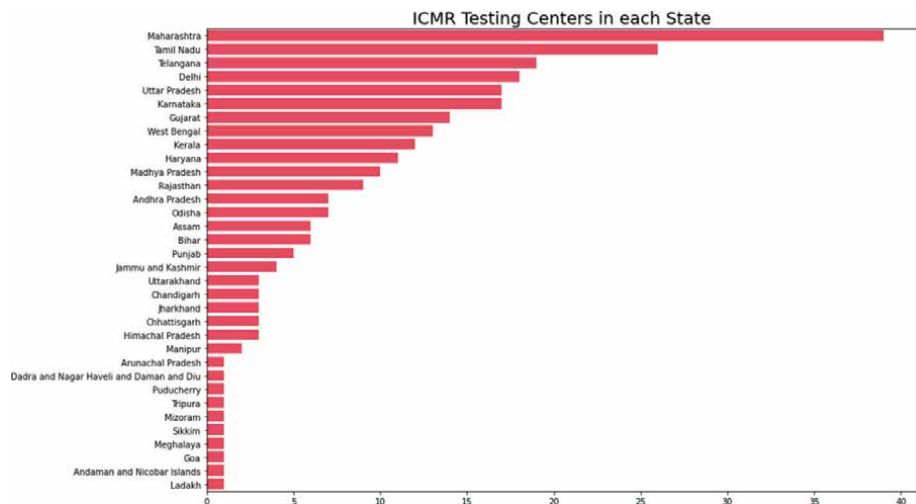


Figure 11.
Distribution of sampling by the Indian Council of Medical Research in different states.

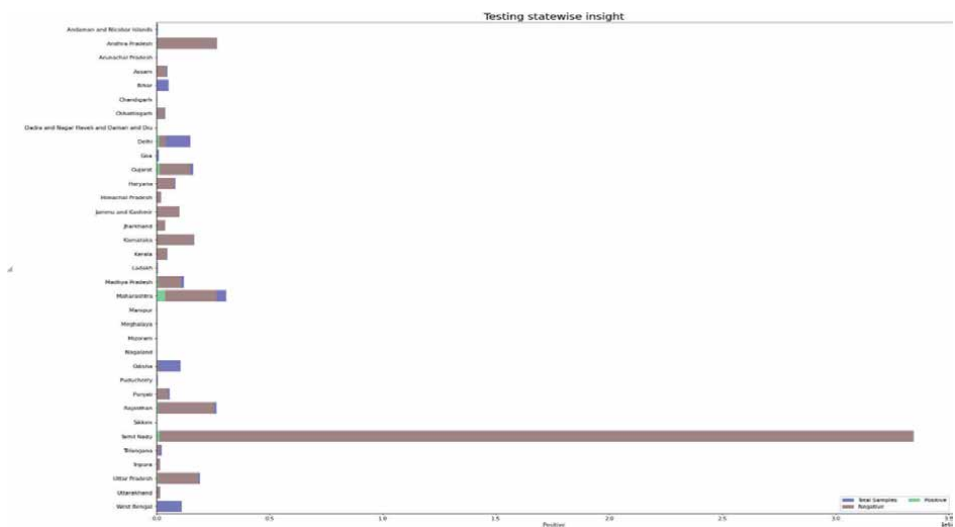


Figure 12.
Visualization of testing ditribution per stat.E.

4.4 PROPHET and ARIMA prediction of the spread in India

Prophet provides fast, reliable and tunable predictions. By developing a base model without adjusting any seasonality-related parameters or adding any extra regressors, it was possible to predict confirmed cases of COVID-19 2 weeks in advance using Prophet and to estimate the number of deaths each day with a 95% prediction interval. By dialing the Prophet, it was possible to plan the forecast. Passing in your forecast data frame and plotting. **Table 2** is showing the produced prediction of total deaths per day based on daily reported figures. Starting from May 24, 2020 to June 7, it was possible to use the previous scenarios to predict. As it is shown in the prediction

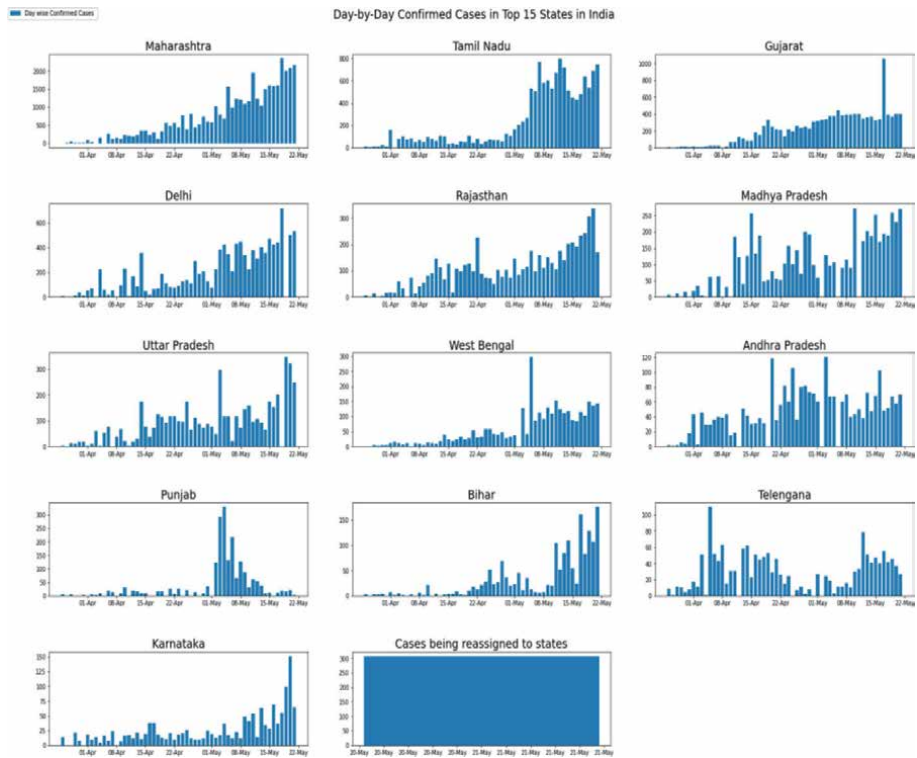


Figure 13.
Top 15 Indian states with confirmed COVID-19 between April and may 24, 2020.

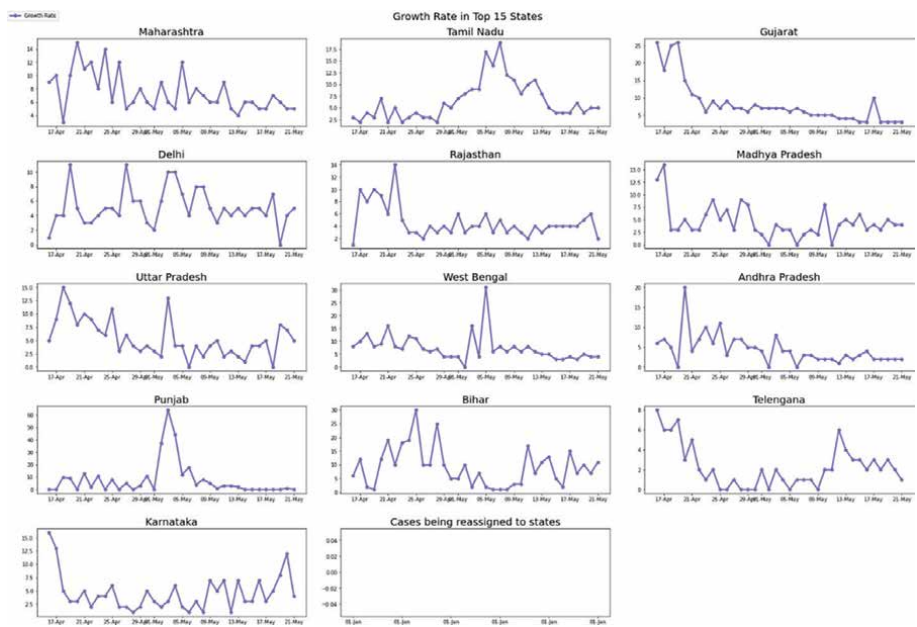


Figure 14.
Top 15 Indian states with growth rates in confirmed cases between April and may 2020.

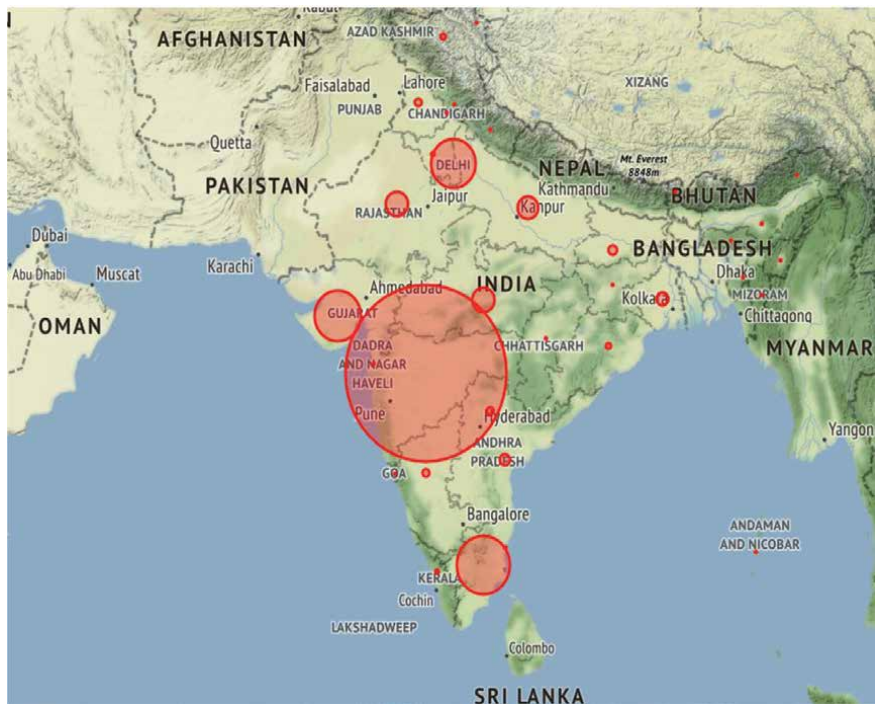


Figure 15.
Geospatial visualization of hotspots of the spread in India as of may, 24, 2020.

Deaths	Days
123	2020-05-24
124	2020-05-25
125	2020-05-26
126	2020-05-27
127	2020-05-28
128	2020-05-29
129	2020-05-30
130	2020-05-31
131	2020-06-01
132	2020-06-02
133	2020-06-03
134	2020-06-04
135	2020-06-05
136	2020-06-06
137	2020-06-07

Table 2.
Mplementation of Prophet prediction between may 22 and June 7, 2002.

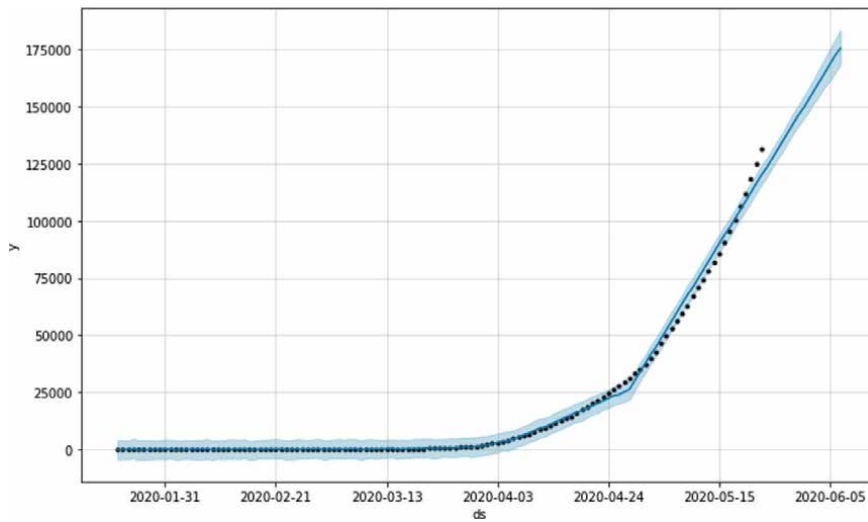


Figure 16.
15 days PROPHET prediction with 95% interval.

model, it was found that the death toll was increasing, given the historical number and was expected that it will reach its peak by June 6, 2020, with a total number of 137 deaths. **Figure 16** is showing the plotting of the model along with the actual as it was found there was predicted for 1 week, until June 6, 2020.

The ARIMA model which stands for Auto-Regressive Integrated Moving Average model was used to predict for 15 days, from the last day of reported confirmed cases. The model explains the given reported death cases based on its previous past values to predict future values [16, 17]. Non-seasonal time series, as the one used in this research exhibits a pattern that is not random. The pattern order of the auto-regression, the order of the moving average and the difference between them, is required to make the time series stationary. According to the results, the global pattern of increasing spread and impact of affected reported confirmed cases, as well as the number of deaths, was on the rise for the period of the study. However, India has a special pattern that the spread majorly in the age group of less than 49 years old. **Figure 17** is showing the growth pattern of the confirmed cases as modeled by ARIMA.

4.5 India versus the global impact of COVID-19

The response of India to COVID-19 was quick and efficient. A proactive nation-wide lockdown as place to help with flattening the curve and to enhance the resources and capacities to deal with the response adequately. Comparing to the global impact the growth of reported cases was growing exponentially. **Figure 18** is showing the increase in total cases between China, Germany, Iran, Italy, Spain, the United States of America, South Korea, France, the United Kingdom, and India. It is noticeable that the rise of cases in India in millions, between early February and mid-May, 2020. The large population can be a justifying factor. However, the number of tests is reported to be limited, which may have led to an increase in reported cases. **Figure 19** provides visualization of the global statistics of the recorded, active, and deceased cases for the end of the study period. It was found that 59.1% of cases were active cases, compared to recovered 35.2% and 5.7% decreased.

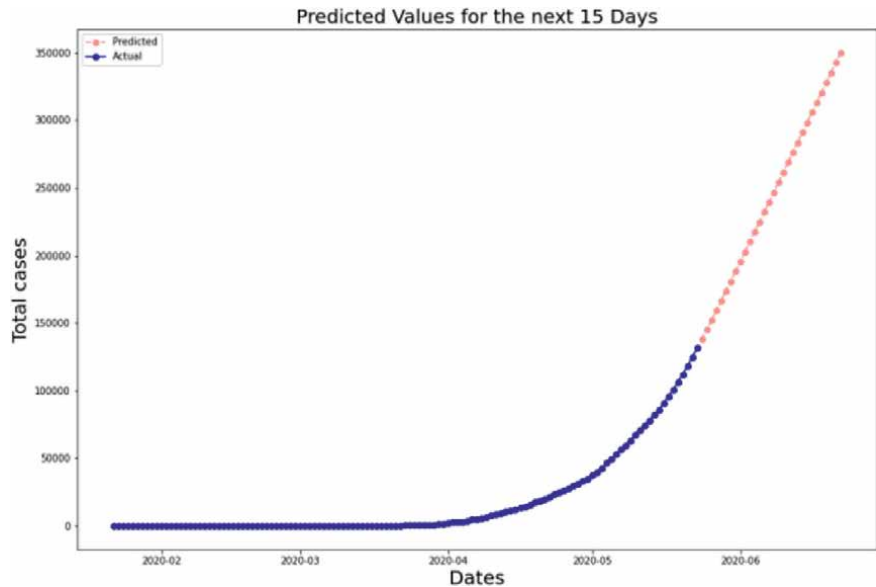


Figure 17.
15 days, ARIMA prediction of the total cases starting from may 24, 2020.

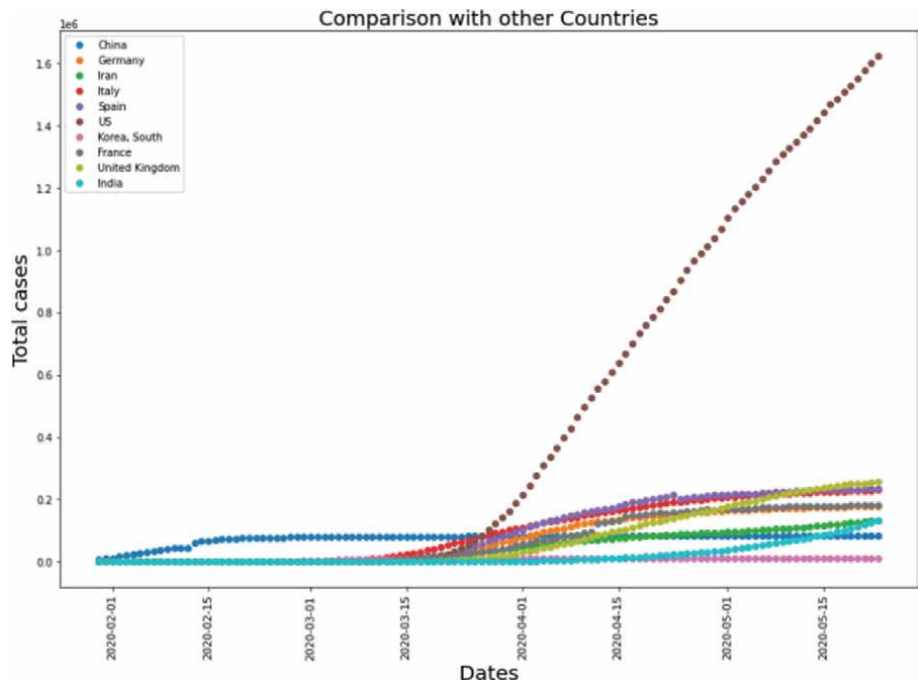


Figure 18.
The severity of opread of Indian versus other countries' reported cases.

Figure 20 provides a visualization of global hotspots with active, recovered and deceased cases in China, Germany, Iran, Italy, Spain, the United States of America, South Korea, Turkey, France United Kingdom and India, suggesting that the total active cases in India were 120,000 case, by the end of May and the rise have begun

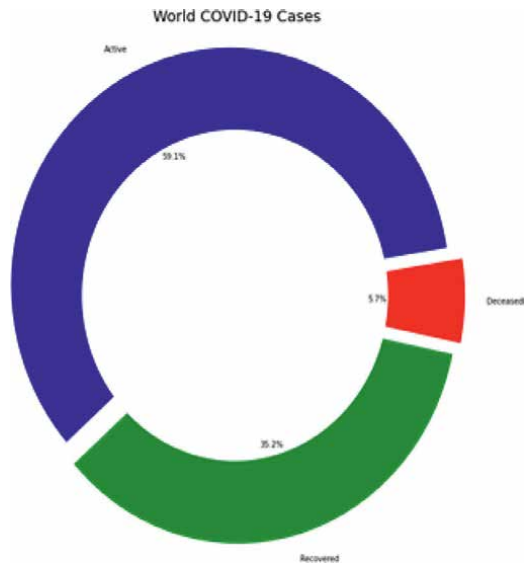


Figure 19.
Recovered, active, and deceased population of COVID-19 as of May 24, 2020.

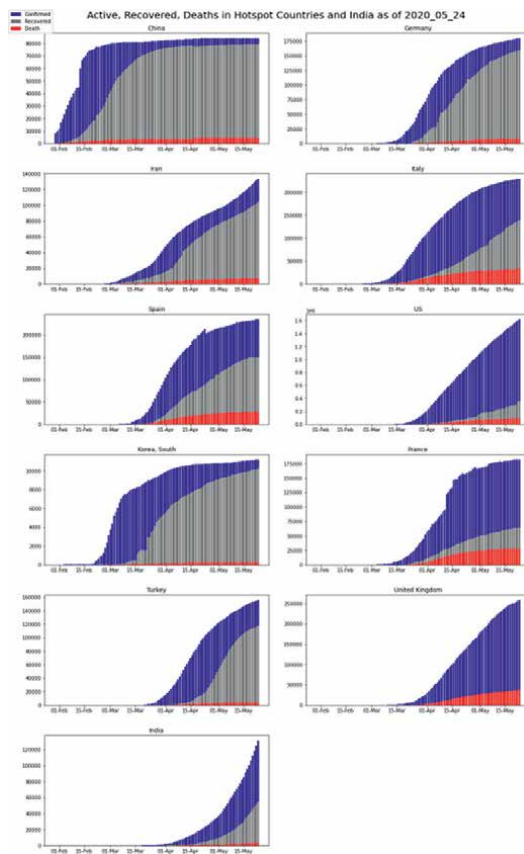


Figure 20.
Comparison between different hotspots of active, recovered, and deaths as of May 24, 2020.

from early April. However, for China, the is showing that from Feb till May the total active cases were at 80000 cases. Germany has witnessed 75,000 between April and May, which brings Germany to one of the hottest spots in the spread of COVID-19.

5. Conclusions

Visualization of complex and big data for delineating spatiotemporal patterns for disaster management and emergency response has been considered widely during the recent spread of COVID-19. Visualizing invisible patterns of data and spatially enabling decision-makers with information contributes to the strength of knowledge that is mostly needed in extreme situations. Sharing data and allowing various contributing domains freely and interoperable access big data of COVID-19 provide added value for medical resources optimization, planning and responding, patterns recognition, and the overall situational awareness that is highly demanded at the planning level, operational level and decision-making levels during pandemic disasters. Visualization of spread in terms of active cases, recovered and deaths showed that India has drastically surpassed Germany, Iran, Italy, Spain, the United States of America, South Korea, Turkey, France, the United Kingdom began starting early April 2020.

This research showed that India as a highly populated country have the capability to leverage technological tools and products in response to extreme situations, through sharing, accessing, collaborating and integrating various decision-making models for rapid and effective response as to how and why scientific visualization and machine learning modeling can be effective in pandemic situations. Prophet Model and ARIMA model were used to predict for 2 weeks using the historical reported data prior to the prediction. The results of this research confirm that Machine Learning visualization and prediction are effective in disaster management situations, dealing with large outbreaks. It showcases the national efforts in specific countries and correlates with the international data. Visualization of COVID-19 spread on different demographic groups showed resilience in the group over 30 years old and over 59 years old, resulting in an increasing and forecasting a more promising growth of population.

Acknowledgements

The authors would like to acknowledge Health Care authorities of India, for making their data and API accessible for research.

Conflicts of interest

The author declares no conflict of interest.

Author details

Rifaat Abdalla^{1*} and Imen Hamdi Nasr²

1 Department of Earth Science, College of Science, Sultan Qaboos University, Oman

2 Department of Geology and Geophysics, College of Science, University of Science of Bizerte, Tunisia

*Address all correspondence to: rabdalla@squ.edu.om

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Mishra K, Rampal J. The COVID-19 pandemic and food insecurity: A viewpoint on India. *World Development*. 2020;**135**:1-3. DOI: 10.1016/j.worlddev.2020.105068
- [2] Andrews MA, Areekal B, Rajesh KR, Krishnan J, Suryakala R, Krishnan B, et al. First confirmed case of COVID-19 infection in India: A case report. *The Indian Journal of Medical Research*. 2020;**151**:490-492
- [3] Narayanan L, Pandit M, Basu S, Karmakar A, Bidhan V, Kumar H, et al. Impact of lockdown due to COVID-19 outbreak: Lifestyle changes and public health concerns in India. *Preprints*. 2020;**1**:2020060129. DOI: 10.20944/preprints202006.0129
- [4] Hongyan Ren L, Zhao AZ, Song L, Liao Y, Weili L, Cui C. Early forecasting of the potential risk zones of COVID-19 in China's megacities. *Science of the Total Environment*. 2020;**729**:1-8. DOI: 10.1016/j.scitotenv.2020.138995
- [5] Pourghasemi HR, Pouyan S, Heidari B, Farajzadeh Z, Shamsi SRF, Babaei S, et al. Spatial modeling, risk mapping, change detection, and outbreak trend analysis of coronavirus (COVID-19) in Iran (days between February 19 and June 14, 2020). *International Journal of Infectious Diseases*. 2020;**98**:90-108, ISSN 1201-9712. DOI: 10.1016/j.ijid.2020.06.058
- [6] Aljaaf A, Hussain A, Fergus P, Przybyla A, Barton GJ. Evaluation of machine learning methods to predict knee loading from the movement of body segments [Internet]. In: *Neural Networks (IJCNN)*. IEEE World Congress on Computational Intelligence, Canada, 24-29 July 2016
- [7] Brynjolfsson E, Mitchell T. What can machine learning do? Workforce implications. *Science*. 2017;**358**(6370):1530-1534
- [8] Utkin V, Zhuk YA. An one-class classification support vector machine model by interval-valued training data. *Knowledge-Based Systems*. 2017;**120**:43-56
- [9] Khan FN, Zhong K, Al-Arashi WH, Yu C, Lu C, Lau APT. Modulation format identification in coherent receivers using deep machine learning. *IEEE Photonics Technology Letters*. 2016;**28**(17):1886-1889
- [10] Polese M, Jana R, Kounev V, Zhang K, Deb S, Zorzi M. Machine learning at the edge: A data-driven architecture with applications to 5G cellular networks. In: *IEEE Transactions on Mobile Computing*. 2018. DOI: 10.1109/TMC.2020.2999852
- [11] Nijeweme-d'Hollosy WO, van Velsen LS, Poel M, Groothuis-Oudshoorn CG, Soer R, Hermens H. Evaluation of three machine learning models for self-referral decision support on low back pain in primary care. *International Journal of Medical Informatics*. 2018;**110**:31-41
- [12] Xie Z, Qin Y, Li Y, Shen W, Zheng Z, Liu S. Spatial and temporal differentiation of COVID-19 epidemic spread in mainland China and its influencing factors. *Science of the Total Environment*. 2020;**744**:140929. DOI: 10.1016/j.scitotenv.2020.140929
- [13] Zhou C, Fenzhen S, Pei T, Zhang A, Yunyan D, Luo B, et al. COVID-19: Challenges to GIS with big data. *Geography and Sustainability*. 2020;**1**(1):77-87. DOI: 10.1016/j.geosus.2020.03.005

[14] Italian Civil Protection Department, Morettini M, Sbröllini A, Marcantoni I, Burattini L. COVID-19 in Italy: Dataset of the Italian civil protection department. Data in Brief. 2020;**30**:1-7. DOI: 10.1016/j.dib.2020.105526

[15] Franch-Pardo I, Napoletano BM, Rosete-Verges F, Billa L. Spatial analysis and GIS in the study of COVID-19. A review. Science of The Total Environment. 2020;**739**:1-10. DOI: 10.1016/j.scitotenv.2020.140033

[16] Lary DJ, et al. Machine learning applications for earth observation. In: Mathieu PP, Aubrecht C, editors. Earth Observation Open Science and Innovation. ISSI Scientific Report Series. Vol. 15. Cham: Springer; 2018. DOI: 10.1007/978-3-319-65633-5_8

[17] Khan FM, Gupta R. ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. Journal of Safety Science and Resilience. 2020;**1**(1):12-18. DOI: 10.1016/j.jnlssr.2020.06.007

The Synergy of Data Science, GIS Spatial Analysis and Knowledge Management as a Path to Sustainability Insights

Mohamed B.O. Osman

Abstract

This chapter focuses on data science, GIS-based spatial analysis, and knowledge management (KM) threesome and its potential contributions to sustainability insights. Discussion commenced with tracing the historical evolution and essentiality of geography for comprehending environmental and social dimensions of sustainability. Then, argumentation delved into the interplay among the three-some's domains and their contributions to sustainability achievement. A prolonged elaboration was provided on geospatial data analytics, visualization, geospatial data mining, and predictive models and their significance for extracting informative and meaningful insights. Since data science has transformed and enriched most scientific disciplines, its empowering implications on GIS were explained. Spatial analysis, therefore, occupied a central position and enabled advanced GIS technique utilization to reveal patterns, relationships, and trends in geospatial data. Furthermore, the chapter explained the interdependent relationships between GIS and KM. Integrating GIS and KM techniques has revolutionized geospatial data organization, visualization, and dissemination and enhanced applications of decision support, environmental planning, and others. The Nexus of this threesome, therefore, serves as a roadmap for solving issues of intricate spatial problems *via* modeling and informed decisions. The chapter stressed and concluded that the integrated fusion of data science, GIS, and KM provides a robust framework and ideal tools supporting sustainability.

Keywords: informatics, data science, GIS-spatial analysis, KM, sustainability

1. Introduction

Space where humans exist is vital for their lives and activities. Thus, geography has always been of significance to humans. Since times immemorial, with concerns and ongoing endeavors, man has been striving to understand the phenomena and objects in his surrounding environment to enable utilization as well as prevention of hazards and disasters nature holds for the sake of a prosperous and safe life.

In this light, the stone-age hunters were anticipating the locations of the quarry, early explorers' deaths or lives were determined by their geographic knowledge, and contemporary societies would not be able to plan and perform their everyday activities without asking and finding geographic answers such as where and how far, [1].

Over these centuries of lengthy and multiple interactions, humans have known and experienced nature by benefiting from its resources and, at times, being traumatized by the disasters it causes. This, besides developing approaches for spatial thinking, had accumulated tremendous knowledge about the surrounding objects and phenomena and composed the scientific heritage, especially in various fields of the so-called "life and applied sciences", such as geography.

Thanks to these experiences as well as the spatial thinking approach, humans have accumulated spatial knowledge that gradually varies over time and space, on the one hand, and from knowledge about objects or phenomena characteristics to their interrelationships, which all together comprise the surrounding environment, on the other hand.

For this reason, knowledge, especially geographic knowledge, starts with and varies between simple and complicated forms. The simple form is the level of describing the object or phenomenon and identifying its characteristics, while the second form is detecting the interrelationships among these objects or phenomena and understanding how they affect each other's [2].

Historically, and prior to the technological and communication revolution, the sources of knowledge were basically confined to two sources: direct and personal observation or experience and the indirect experiments of others [2]. This caused knowledge accumulation and dissemination over time and space to become very decelerated and limited.

Aftermath, this indolent and gradual knowledge exchange had been abruptly disturbed by iconic events and great breakthroughs, such as the prosperity of the paper industry and the invention of the typing machine. These backboneed and revolutionized various forms of printed press and hence dramatically changed the production and dissemination of information and knowledge, reaching to the recent situations of information and the technological revolution led and shaped by the internet.

This era is, therefore, the era of information and technological revolution, [3] which is forged and directed by the extensive and diversified range of human activities and the general advancements in all aspects of life. Awing to automation, economic growth, urbanization as well as the technological explosion, especially the advancements of communication means and computerized information technology.

These factors had drastically changed and modernized the availability and accessibility of information and knowledge through developing data collection, processing, and dissemination techniques. This dissemination is embodied in the obtainable and massive volumes of knowledge and databases, data warehouses, and big data archives. Abundance of these enormous data, in both levels of phenomenon and inter-phenomenal escorted, consequentially, urbanization and industrialization, besides the ever-growing complexities in different life aspects.

In the same context, the increasing population and their economic activities have urged the dire need to take appropriate decisions regarding development, resources utilization, society, and environmental conservation [4]. Likewise, it aroused the development of techniques to utilize and extract useful and practical-guiding knowledge out of these compilations of knowledge and data bases.

These techniques include, for instance, GIS spatial analysis, spatial decision support systems, and expert systems. All of these have the purposes of enabling us to simplify, model, and abstract the complex and thorny reality for better understanding as well as supporting data-driven and informed decisions for best and friendly practices for man and his surrounding environment alike.

2. A brief epistemological framework of data science and GIS-spatial analysis

As side-effects of the overwhelming digitalization, the massive advancements in computation capabilities and information and communication technologies affected approximately all disciplines and professions. It was thus argued that data science had shifted and redesigned all scientific disciplines, by means of adjusting their adopted approaches [5].

Therefore, new trends and developments in science and research emerged and caused many fields and disciplines to revisit and upgrade their methods and practices to become more IT-oriented [6]. Furthermore, the merging of geography into other fields was noticeable in many disciplines that engaged the spatial dimension in their approaches by being oriented to handle more spatially referenced information and data through the platform of geographic information systems, or GIS. This indicates growing concern, cognition and consideration of geographic and spatial perspectives.

Geography itself, even prior to the computation and information explosions, had reshaped its nature, through adapting newer approaches and transforming from the traditional descriptive geography (regional and systematic geography) to the so-called, applied geography. This added value and dimensions of practicality to geography as a problem-solving scientific discipline.

Applied geography, in the form of maps and spatial information, has served discovery, planning, cooperation, and various management-related practices for not less than 3000 years. For this reason, it was debated that most of the geographic knowledge is applicable to routine tasks [1]. Yet this seems to not be enough and satisfying to the requirements of the state-of-the-art in informatics and modern technological situations in contemporary society.

To further cope with the general advancements in all of life's aspects and the complexities of the rapidly changing world, geography has shifted its paradigms and approaches; thus, it has undergone different phases, which can generally be expressed in the aforementioned transformation from descriptive to applied geography. Of course, the quantitative revolution, digitalization, and engagement of computer graphics were the most important and influential milestones toward that transformation.

These all paved the road and led to the emergence of a newer field and branch of applied geography, the geo-informatics, that endeavors to take advantage of the advancements in information and communication technologies' infrastructure to leverage the various applications of geography. Because spatial information is paramount, we have developed tools and techniques called Geographic Information Systems (GIS) to handle geospatial information and support us with geographic knowledge. GIS enable us to collect and utilize spatial data (See **Figure 1**) [1].

GIS is a computer-based system that facilitates the development and use of spatial data. This is typically known as spatial analysis. There are many reasons for developing, utilizing, and adapting spatial analysis methods and general GISciences, the most

prominent factors are societal and technological. The societal reason is basically the need to more effectively and efficiently use the resources, alongside consideration of the environment, finite resources, and growing population. While technological reasons are the interest and necessity of applying the new tools and technology to handle these thorny issues.

GIS technology as well as its methods of spatial analysis are therefore based on geographic information science, which is supported and widely applied by various disciplines of geography, surveying, engineering, space science, computer science, cartography, statistics, and others for purposes of spatial data organization, analysis, and distribution to eventually improve life.

Over time, the significance of spatial analysis increased. It utilizes massive amounts of various data formats to serve a wide-range applications of the concerned scientific disciplines. Each discipline uses special software and hardware that suits its mission and data formats. For this, recent practices of GIS and spatial analysis need advanced algorithms and programs to enable proper and timely handling and processing of immense and interdisciplinary data. This resulted in the spatial data science, which is the branch of data science that manipulates the spatial data.

3. Data science essentiality for geospatial big data analytics

Per the previously discussed issue of data inflation, the new discipline of data science was founded to handle and make use of these data floods and ensure getting the maximum out of them. The arena of this newly founded field, therefore, was the data with its different formats. This subsequently resulted in a number of new connected and related concepts and terminologies, such as big data, the internet of things, machine learning, data bases and warehouses, data archives, knowledge bases, expert systems, decision support systems, and many others. Forming the subdivisions

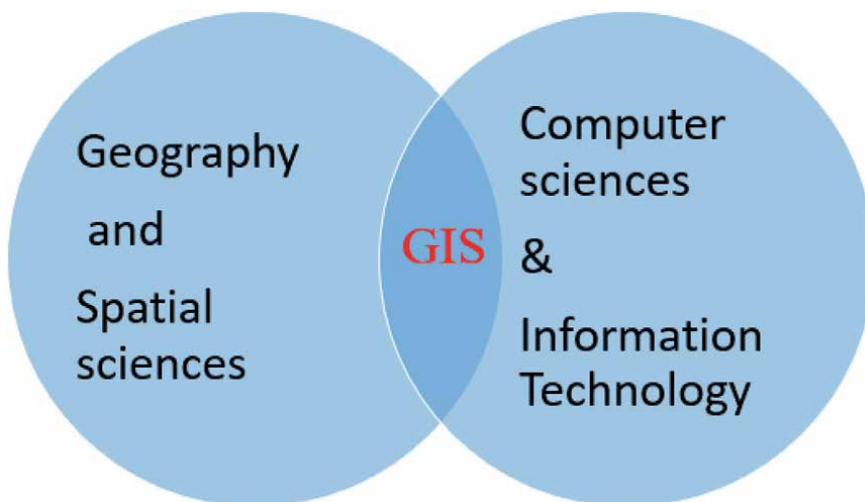


Figure 1.
Integrative nature and position of GIS among other major disciplines. Source: created by the author.

and environment of data science. The below chart, highlights the nested and connected advanced technologies for data analytics (See **Figure 2**).

Big data is one of the prominent beneficial effects and usefulness of data science; nevertheless, it is still not a clearly defined term, and it has been defined differently according to each of the technological, industrial, research, or academic viewpoints [8]. However, in general, big data is considered structured and unstructured datasets with massive data volumes that cannot be easily captured, stored, manipulated, analyzed, managed, and presented by traditional and ordinary hardware, software, and database technologies [9].

The dire needs of analytics as well as newer methods and techniques of data science emerged as a response to the current enhancements in life aspects generally and the quest for better living conditions. Accordingly, humans—via various disciplines—strive to make use of the recent advancements in ICT researches and technologies to cope with reality’s complexities through data and models. For this, big data emerged and prevailed in most of science fields. Geography and geospatial sciences generally have the most stake, as reflected in “geospatial big data”. It was argued that “80% of the data is of a geographic nature.

Despite criticisms regarding the societal effects of big data, which has been accused of having limited benefits for society and violating issues related to data privacy, confidentiality, and security, geospatial big data, nonetheless possesses significant potential to greatly benefit various societal applications. These applications

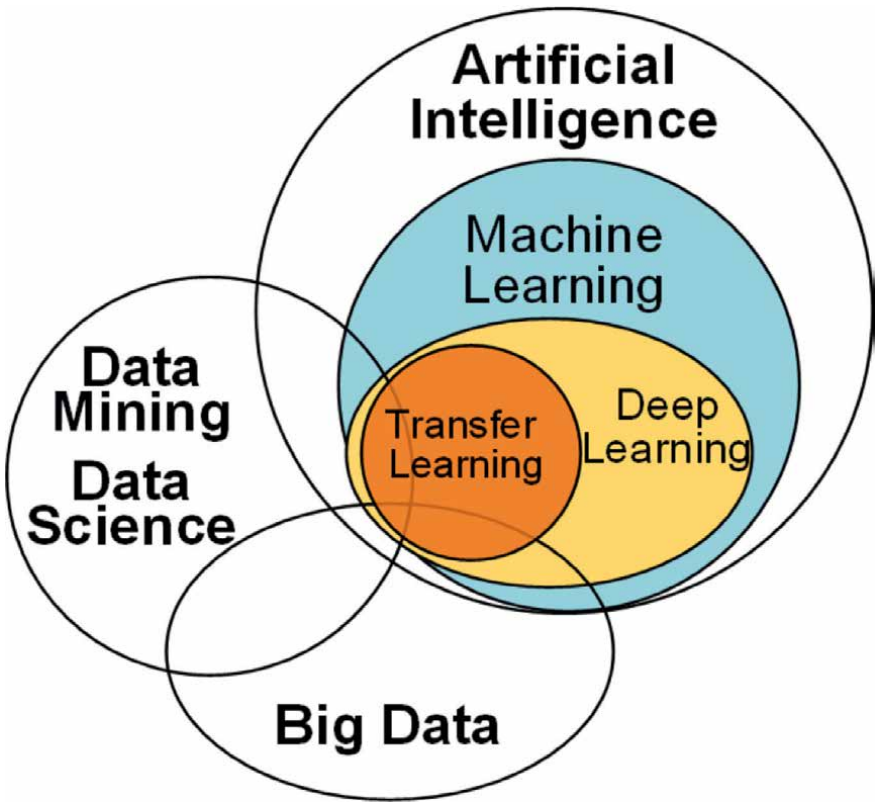


Figure 2.
Demonstrates the interplay of recent data science and analytics techniques. Source: adopted from Li et al. [7].

encompass areas such as climate change, disease surveillance, disaster response, the monitoring of critical infrastructure, transportation, and more [9].

Geospatial big data processing techniques encompass a range of methodologies designed to handle and analyze vast volumes of geospatial data due to programming models and frameworks for distributed processing of geospatial data [10]. These techniques are vital in extracting valuable insights from geospatial data, aiding decision-making in a spectrum of various domains such as urban planning, environmental monitoring, and disaster management.

In order to be handled and processed, the available big data volumes need to be compressed without compromising data quality [8]. The parallel processing and distributed computing frameworks, therefore, came and leveraged the computational demands of geospatial big data [11]. These all go beyond the capacity of the human mind to handle.

Reading and understanding big data patterns and trends—the minimal level of benefits—is not possible without the aid of computerized methods of summarization and visualization. Further benefits could be yielded by deploying knowledge discovery algorithms and data mining techniques for the extraction of more valuable and well-oriented knowledge that serves problem solving, decision support, and even strategic policy-making on the basis of deep understanding.

In the same context, visualizing this variable and complex data needs a computerized and graphical capability to facilitate communication of these data. The data in its real forms and row formats is not understandable and readable, and both forecasting future scenarios and making the right decisions for the present based on the available data are not possible without developing algorithms and complicated models. The human mind cannot handle the complicated calculations for the big data volumes. This necessitated and urged recent data science to embody.

4. Spatial analysis in the information era: the emergence of spatial data science

The ultimate goal of science, the applied and life sciences in particular, is to support humankind and equip it with the facilities needed for a prosperous life. Science, thus, augments sustainability by employing advanced analytics and machine learning techniques to extract actionable insights from vast and diverse geospatial datasets. This analytical ability enables proactive identification of environmental trends, prediction of risks, and optimization of resource allocation strategies [12].

Historically, observation—the direct field observation—was the sole prevailing source of data, and decision-making was not complicated, so the human mind was capable of handling that. However, the informatics insurgency came to digitalize and automate most activities and life aspects. In science, data collection, processing, storage, and visualization had been profoundly reformulated by computerization and informatics, which led to the information era.

In the Information era, spatial analysis has emerged as a crucial discipline, facilitated by data availability and advancements in technology. Spatial analysis further encapsulates this transformative trend and highlights the significant role of spatial data science in extracting insights from vast and complex geospatial datasets. Therefore, emergence of spatial analysis and geospatial sciences is featured by their capacity to address multifaceted challenges across diverse domains. See the below chart, **Figure 3**. It demonstrates the spatial data science emergence.

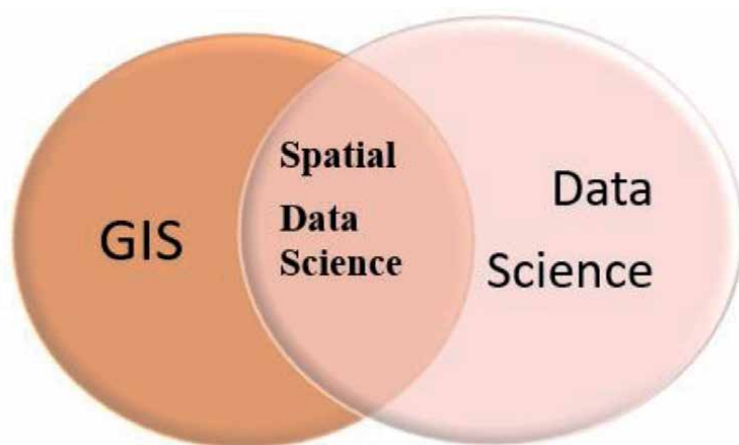


Figure 3.
Emergence of spatial data science due to integration of GIS into data science. Source: created by the author.

In its generic and plain definition, spatial analysis refers to the process of examining, modeling, and interpreting patterns and relationships within geographic data to gain insights and make informed decisions about the spatial distribution of phenomena. Spatial analysis further encompasses a wide range of techniques and methods used to analyze spatial data, including mapping, clustering, interpolation, and geostatistics [13].

The overall goal of spatial analysis and its processes can be categorized into two levels of ultimate knowledge generation. These levels vary between the identification and classification of objects and phenomena, revealing patterns as well as understanding the relationships among these phenomena—how objects interact and affect each other. The spatial analysis, thus, strives to model and abstract reality, particularly what regards geography and has spatial dimension, to enhance understanding and, subsequently, enable better decision-making and problem solving.

Thus, the complex reality and synergy of geospatial technologies, data science methodologies, and advanced analytics insisted and compelled a deeper understanding of spatial relationships, patterns, and trends, and was the driver the paradigm shift in spatial analysis. This caused the spatial analysis to play a crucial role in various fields such as geography, environmental science, urban planning, and epidemiology, [14].

Furthermore, due to its ability to reveal hidden patterns and trends that may not be detectable through the traditional statistical analysis, spatial analysis gained special interest among researchers and practitioners. This novel approach thus become inevitable in sophisticated applications of natural resources management, urban planning, environmental monitoring, epidemiology, and all spatial-related fields. This is what enabled it to drive evidence-based decision-making and contributing to more sustainable and insightful policies [11].

In the contemporary context, the intricacies of life are induced by the fusion and interplay of phenomena, exemplified, for instance, in the nexus of environment, economy and society besides, the global landscape characterized by swift transformations and changes, wherein the accumulated data and knowledge are combined with complicated decision-making processes. These processes, in their multifaceted nature, require the consideration of numerous dimensions simultaneously.

These facts necessitate reading the massive data volumes, of different sources and formats, available in databases, data warehouses, and obtainable knowledge bases to fully understand the current state of affairs. Spatial indexing and query optimization would be the right and ideal techniques for that, as they enable organizing and handling various data besides creating efficient spatial indexing structures and query optimization methods to accelerate spatial queries [15].

Geospatial Data Mining, in the same context—taking advantage of the immense availability of integrated and multi-sourced data—emerged to apply data mining and machine learning techniques to extract patterns and relationships from geospatial datasets [16, 17]. Especially these data, in order to give meaningful values, should be subjected to careful and rigorous analysis through well-oriented questions that extract insights and valuable knowledge for problem solving and decision-making out of those massive data volumes. Geospatial Data Mining provides an effective, efficient, and fast mechanism to do so.

5. Common data science processes and approaches in GIS context

The application of data science to different disciplines is inevitable in the era of technology and informatics. It was previously elaborated how data science had practically affected the various scientific fields. Geography is not an exception and had been one of those affected the most. This, along with the recent merger of geography and data science resulted in the newly found branch of geography, known as geographic information systems. Some scholars went beyond that and perceived GIS as an independent discipline “geospatial data science”.

Definitions of GIS unanimously agreed that GIS is basically the science of spatial data collection, processing, and visualization. All of GIS’s functions have thus been deeply altered by data science. Data science has a number of manifestations in the GIS world. For instance, transforming the way spatial data collected and interpreted alongside the inclusion of spectrum of methodologies that leveraged GIS analytics, machine learning, and data mining techniques to extract valuable insights from geospatial datasets.

Prior delving into the details manifesting data science role in GIS world, the following diagram gives a quick glance and outline the GIS and data science syntheses, by means of highlighting the techniques and approaches of data science GIS had adopted and their respective applications and uses (See **Figure 4**).

In the GIS world, data preprocessing is a foundational step involving data cleaning, transformation, and integration to ensure the quality and consistency of geospatial datasets [16, 17]. This aligns with data science’s emphasis on data preparation for meaningful analysis.

The infusion of data science processes and approaches into the GIS context has ushered in a new era of spatial analysis and understanding. Data fusion and integration are among essential data science processes in the GIS context. They allow combining data from diverse sources, such as satellite imagery, maps, and other forms of statistical and descriptive data, from which analysts can create enriched datasets that offer comprehensive insights into various phenomena [18]. This is particularly valuable in disaster management, where the integration of “nearly” real-time data streams aids in tracking, monitoring and responding to natural hazards.

In recent years, along with the availability of new sensors, new ways of collecting geospatial data have emerged [7], which means the digital data collection methods

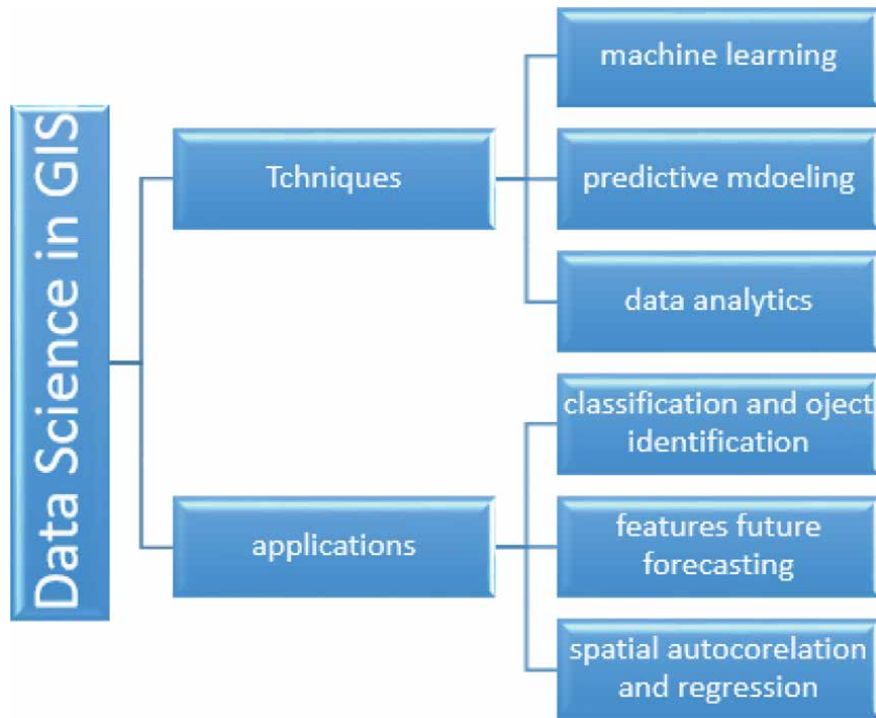


Figure 4.
Highlights some prominent data science approaches, techniques and tools adopted in GIS. Source: created by the author.

introduced and accompanied by data science have shifted data collection and transformed it from sparse data to rich and plentiful databases. This situation had been constantly developing until the introduction of spatial data infrastructure (SDI). Data collection, likewise, had evolved and reached the recent state of availability of nearly real-time data and, at certain times, real-time data.

Awing to SDI and massive existing geospatial data, geospatial data exploration and visualization which is one of core data science practices found resonance within GIS. Interactive maps, spatial dashboards, and geo-visualizations provide insights into spatial patterns and relationships, aiding decision-making and the communication of complex information to all different audiences [19]. Furthermore, spatial clustering and pattern recognition, are key techniques in data science that are extensively deployed in GIS analytics for identifying hotspots, spatial trends, and clusters in geospatial data. (See **Figure 5**) [20].

Moreover, spatial autocorrelation and spatial regression techniques, such as geographically weighted regression, have gained prominence in both data science and GIS. In GIS, they allow the identification of spatially varying relationships among given variables, enhancing predictive accuracy in local contexts [13]. Another form of spatial clustering is the process of grouping spatial data points based on their similarity. Methods like K-Means clustering are applied in GIS to fragment geographic data into clusters, revealing spatial patterns and helping to identify regions with similar characteristics [22]. This has frequent use in urban planning, where it aids in delineating distinct neighborhoods based on their common socioeconomic attributes.

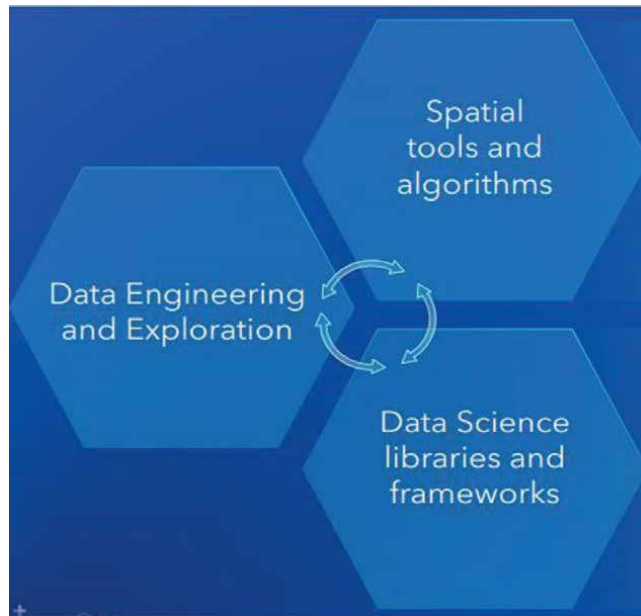


Figure 5.
Elaborates how do some data science tools deployed and operated by GIS. Source: adopted from [21].

Deep learning methodologies and Machine learning algorithms developed by data science have leveraged a wide range of GIS processes. Deep learning tools, such as Convolutional Neural Networks (CNNs), have enriched object detection and image classification within GIS domain. CNNs have proven to be effective in identifying objects and features in satellite images [23]. While machine learning facilitated predictive modeling and image classification tasks. Supervised and unsupervised machine learning methods grew in popularity as a frequent tool for land cover classification, species distribution modeling, and urban growth prediction [24]. See the below diagram.

One of the key machine learning algorithms for spatial prediction and classification is Random forest. Random forest enables the creation of predictive models that account for spatial relationships, allowing for the identification of trends and patterns in complex geographical data [12]. For instance, ecologists, in order to model species distribution, utilize machine learning to foresee the spatial distribution of species based on certain environmental variables. Likewise, Machine learning algorithms like random forest and support vector machines leverage spatial data layers to predict outcomes like land cover changes or disease outbreaks [25].

Therefore, by harnessing the power of data science, GIS practitioners can generate accurate predictions and enhance their understanding of complex spatial patterns. Wherein large and diverse datasets are processed and analyzed to extract meaningful insights. This involves techniques like data clustering and spatial interpolation, which facilitate the extraction of valuable knowledge from vast amounts of geospatial information [26].

Regarding the institutional part of GIS, the integration of data science and GIS is also evident in spatial optimization problems. Likewise, location-allocation models, the classic GIS application, had important benefits from data science optimization algorithms to determine optimal facility locations and resource allocations [27].

This fusion of practices and the interdisciplinary approach enhanced the efficiency and accuracy of GIS as decision support systems.

Integration of data science processes and approaches has, consequently, become increasingly prevalent in GIS practices, ranging from data collection and preprocessing up to the final stage of laying out the results. These all have led to enhanced understanding and utilization of spatial information. Hence, a transformative shift in the capabilities of GIS enabled more sophisticated analysis, modeling, and visualization of geospatial information.

6. GIS, KM and data science: the indispensable sustainability nexus

The relationship between Geographic Information Systems (GIS) and Knowledge Management (KM) is interdependent, with GIS serving as a powerful tool to facilitate various aspects of knowledge management. And KM enhances the management and utilization of geospatial information, fostering a structured approach to acquiring, storing, sharing, and utilizing spatial knowledge. This interplay enhances the accessibility, organization, and utilization of geospatial information within both GIS and knowledge management frameworks. The following diagrams explain the main functions and processes of knowledge management and their cyclic nature respectively (See **Figures 6** and **7**).



Figure 6.
Functions carried out in knowledge management. Source: [28].

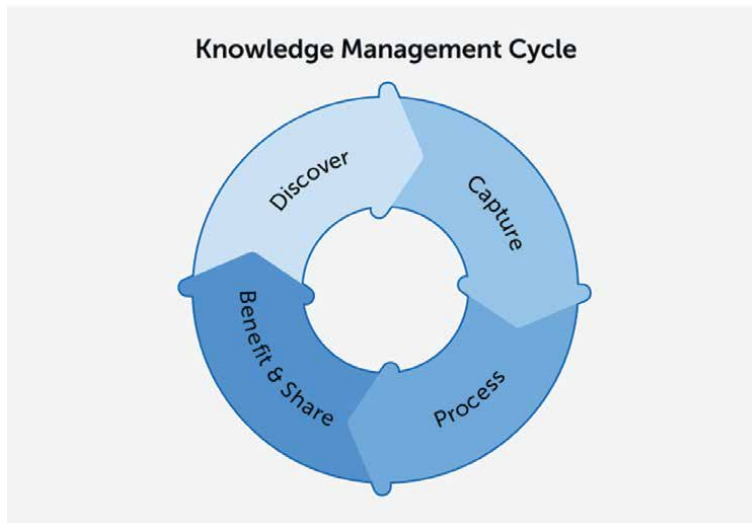


Figure 7.
Demonstrates the cycle of knowledge management. Source: Adopted from [29].

Knowledge management prospects, therefore, are obvious in the contemporary world of GIS. GIS and Knowledge management have recently had a two-way connection; on the one hand, GIS borrowed approaches and techniques from KM for its practices as well as taking the applications and uses of KM and applying them to spatial data. On the other hand, KM takes advantage of GIS as a platform to integrate and utilize multi-sourced data in different formats, all for one unified and oriented purpose. This is encapsulated and briefed in the below table, then the details later (See **Table 1**).

The most prominent knowledge management methods and techniques adopted by GIS encompass a variety of approaches that, together, aim at effectively capturing, organizing, and utilizing geospatial information. To mention and elaborate on some:

1. Geospatial data integration that makes use of knowledge management practices and guides the categorization, tagging, and metadata documentation of this data for effective data organization. KM allows GIS to put together and integrate various geospatial data, for instance, spatial and attributive data, from diverse sources and formats to create an inclusive and coherent knowledge base [30]. This will enable comprehensive spatial analyses, allow users to consider a wide range of variables as well as facilitate well-oriented goals and problem-solving out of them.
2. Geospatial data cataloging and metadata management, i.e. implementing meta-data protocols would lead to standardized documentation, hence easier management of geospatial data, and further enhance data discoverability and usability [31]. In other words, easier and faster running of queries. Thus, efficient meta-data management and documentation ensure geospatial data's proper utilization along with enhancing data discoverability, reusability, and all other aspects of data quality.

No.	GIS Tasks & Processes	Examples for applications	Counterpart terminology	
			In data Science	In KM
1	Data collection and input	Data entry, digitization, different sensors and GPS gatherings	Data collection	Data acquisition
2	Data preprocessing	Image preprocessing, formats unifying, data conversions	Data format conversions and interoperability	Metadata management, data documentation
3	Data storage	SDIs, geodatabases, geospatial big data management	Big data and data bases and archives	Knowledge repositories and warehouses
4	Spatial Data Mining	Geographically weighted regression and identifying patterns in data	Geospatial Data Mining, Pattern Recognition	Knowledge discovery
5	Concurrent data management	Multi-users and enterprise geodatabases	Cloud and parallel computing, DBMSs and	Updates and data integrity
6	Data visualization and summarization	Mapping, geo visualization	Interactive maps	Storytelling, social media
7	Decision support tools	Multi criteria evaluation, suitability analysis, location allocation and location based services	Simulation, scenario and case analysis	Decision support systems and expert systems
8	Spatial relationships extraction	Spatial autocorrelation and regressions, interpolation, hotspot analysis, query	Regression, liner modeling and trend analysis	Time series analysis
9	Modeling	Spatial data models,	Descriptive and predictive modeling	Trends and pattern extraction
10	Data dissemination	Web GIS, interactive and online maps, geoportals	Dashboards, portals an	Knowledge dissemination

Source: created by the author.

Table 1.
Represents some aspects of interchangeability of GIS, KM, data science nexus.

3. Visualization and spatial insights, GIS —thanks to its super powerful tools, like interactive maps —provides visual representations of spatial data, making complex information more understandable. These visualizations put more than one variable together, which aids comparability and enables knowledge managers and stakeholders to grasp spatial relationships, patterns, and trends.
4. Spatial analysis documentation, as knowledge management is a tremendous tool to capture analytical processes. This enables GIS users to replicate spatial analyses, ensuring transparency and facilitating knowledge sharing among researchers.
5. Spatial data infrastructures (SDIs), establishing SDIs would enable enforcing data quality measures while handling and interacting with data. It further facilitates seamless sharing, discovery, and access to geospatial data across

organizations and sectors [32]. Additionally, SDIs combat redundancy in data and result in efficient storage control.

6. Development of geoportal and web-linked data, the main pillar of nowadays web GIS is creating and operating geoportals as centralized platforms for accessing and sharing geospatial information to foster collaboration and knowledge dissemination [33]. Semantically-related data sharing is similarly a web-based. Deploying web-based semantic data sharing technologies revolutionized geospatial data distribution and enhanced the interoperability of and among geospatial data [34].

Knowledge management applications and uses borrowed and adopted to serve in the environment of GIS involve processes and techniques leveraging GIS to achieve its ultimate goals through effectively utilizing the geospatial information. These involved key applications and prominent uses include, among many:

1. Decision support: GIS-based knowledge management aids decision-making by providing spatial insights into knowledge management strategies. KM processes guide the identification of relevant data layers and spatial variables for informed decision support. Through allowing stakeholders to analyze and visualize data to make informed choices in fields like urban planning and disaster management, [35].
2. Collaborative mapping: collaborative platforms offered by GIS enable users to contribute and share data and insights, fostering crowd-sourced mapping efforts for various purposes like community engagement and environmental monitoring. This is obvious when taking advantage of parallel computing. GIS platforms offer collaboration by allowing multiple users to access, process, and share geospatial data and analyses simultaneously. While KM principles facilitate the creation of workflows and protocols for collaborative data sharing, [36].
3. Scenario planning: GIS-driven knowledge management aids in scenario planning by modeling different spatial outcomes and assisting in policy development and resource allocation. KM strategies and workflows help capturing the rationale behind these models, documenting assumptions and methodologies for future reference and analysis.
4. Environmental monitoring: GIS utilizes methods and frameworks originated by the knowledge management discipline to support environmental management through constant environmental monitoring, enabling authorities and decision-makers to monitor and track changes, reveal trends over time, and make informed conservation decisions.
5. Emergency response: Knowledge management in a GIS environment assists emergency response in real-time by providing up-to-date geospatial data for incident management. The synthesis of GIS and knowledge management would also enable the creation and proper operation of early warning systems and empower preparedness for emergencies.
6. Spatial query and retrieval: GIS enables spatial queries for data retrieval. KM techniques assist in designing efficient query systems and indexing approaches for quick and efficient access to relevant geospatial information.

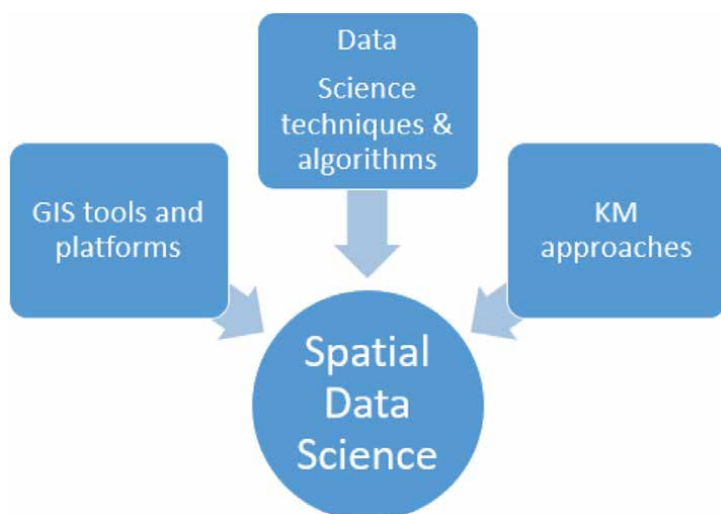


Figure 8.

Explains the synergic threesome of GIS, KM and data science and emergence of spatial data science. Source: Drawn by the author.

For these reasons, one can state that the integration of geographic information systems (GIS), knowledge management (KM), and Data Science has given rise to a transformative paradigm in the realm of sustainability. This three-way relationship represents an indispensable nexus that offers holistic solutions to complex environmental and societal challenges (See **Figure 8**).

7. Conclusion and future directions

This chapter was founded on demonstration and argumentation crowded of cited evidences in order to prove and support of the argument “synergy of data Science, GIS spatial analysis, and knowledge management is a path toward sustainability insights”. In a world characterized by the exponential data growth due to the rapidly changing states and interplay phenomena. GIS, KM and data science nexus emerged to enable establishing of groundworks for understanding those interconnect and shared attributes, and offer a transformative approach for unraveling complex sustainability challenges and the need for informed decision-making.

To Highlight these, the chapter dived into the evolution of data accumulation from historical milestones to the modern digital era. Then it probed the epistemological foundations of Data Science and its profound influence on various scientific domains. Moreover, strived to uncover the manifold manifestations of data science processes approaches deployed in GIS practices ranging from data collection, preprocessing, statistical analysis, to machine learning algorithms and visualization techniques, each one has been examined in the context of sustainability research.

Being in a current world of informatics, essentiality of data science in the realm of geospatial big data analytics, likewise, had been explored in the context and roles of computational analysis in harnessing vast data for sustainability sake. This in response, subsequently, led the emergence of spatial data science field and its

contributions to understanding spatial relationships and patterns. Through, their tremendous techniques that enhance comprehension, modeling, and prediction.

Concluding with the syntheses of GIS, knowledge management, and data science forms an inevitable nexus for achieving sustainability agendas. This integration of geospatial sciences and technologies, enabled data organization, knowledge representation, and collaborative decision-making is explored in the context of sustainability insights.

Author details

Mohamed B.O. Osman^{1,2}

1 Institute of Geomatics and Civil Engineering, University of Sopron, Sopron, Hungary

2 GIS Department, University of Khartoum, Khartoum, Sudan

*Address all correspondence to: mohamedbabekir95@gmail.com

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Bolstad P. GIS fundamentals: A first text on geographic information system. In: *Manual of Geospatial Science and Technology*. 5th ed. White Bear Lake, Minnesota, USA: Eider Press; 2016
- [2] Eltayeb G. *Geography, Environment and Development: The Fundamentals and Roles*. Sana'a, Yemen: Dar Elhikma Elymaniya; 1995
- [3] Osman BT. *Geographic Information Systems, GIS*. Khartoum., Sudan: Open University of Sudan; 2006
- [4] Osman BT. Decision making support and spatial analysis in geographical information systems. *Geographic Letters Series*, 58, Kuwait geographic society & Geography Department. Kuwait: Kuwait University; 2003
- [5] Scheider S, Nyamsuren E, Krueger H, Xu H. Why geographic data science is not a science. *Geography Compass*. 2020;**14**(11):1-15. DOI: 10.1111/gec3.12537
- [6] Osman BT. GIS applications in science and engineering research: Case study of the graduate College University of Khartoum. *Sudan Geographical Journal (SGJ)*. 2017;**1**(1):1-2. Khartoum, Sudan
- [7] Li S, Chen J, Liu C. Overview on the development of intelligent methods for mineral resource prediction under the background of geological big data. *Minerals*. 2022;**12**:616. DOI: 10.3390/min12050616
- [8] Chen X, Li J, Yu L, Zhao Y. Compressing and exploring massive GPS trajectory datasets. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2015;**105**:13-24
- [9] Li S, Dragicevic S, Castro FA, Sester M, Winter S, Coltekin A, et al. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2016;**115**:119-133
- [10] Li Z, Li J, Wu F, Tian Z. A MapReduce-based parallel spatial join algorithm for processing big geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2014;**98**:119-132
- [11] Li Z, Li J, Wu F, Tian Z. Parallel computing for processing massive geospatial data using MapReduce. *Computers, Environment and Urban Systems*. 2017;**64**:225-237
- [12] Gao S, Tang W, Li Q, Yang J. Data science and urban sustainability: A review. *ISPRS International Journal of Geo-Information*. 2019;**8**(1):23
- [13] Fotheringham AS, Brunson C, Charlton M. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester, England: John Wiley & Sons Ltd; 2003
- [14] Openshaw S, Abraham RJ. Geographical information systems and spatial data analysis in the social sciences. *GeoJournal*. 2000;**50**(2):213-224
- [15] Zhang R, Lin H, Liu Z, Shekhar S. G-tree: An efficient and scalable index for spatio-temporal data on multicore processors. *IEEE Transactions on Knowledge and Data Engineering*. 2012;**24**(5):903-917
- [16] Shekhar S, Evans M, Tan X, Zhang H. Advances in geospatial big data handling. *IEEE Transactions on Knowledge and Data Engineering*. 2014;**27**(6):1510-1523
- [17] Shekhar S, Evans M, Tan X, Zhang H. *Advances in Geospatial Big Data Handling: Tools and Methods*. Boca Raton, Florida, United States: CRC Press; 2014

- [18] Mennis J, Hultgren T. Intelligent GIS: Location decisions and strategic planning. In: *The GIS Book*. Boca Raton, Florida, United States: CRC Press; 2006. pp. 245-263
- [19] Khan O, Couclelis H. GIS and the “spatial turn” in information technology. *Geographical Information Science*. 2006;**12**(4):1-11
- [20] Duque JC, Anselin L, Rey SJ. The max-p-regions problem. *Journal of Regional Science*. 2012;**52**(3):397-419
- [21] Kaliski S, Mani A. *How GIS and Machine Learning Work Together*. California, USA: ESRI Redlands; 2019. Available from: <https://www.esri.com/content/dam/esrisites/en-us/about/events/media/UC-2019/technical-workshops/tw-6165-494.pdf>
- [22] Arbia G. *A Primer for Spatial Econometrics: With Applications in R*. London, UK: Palgrave Macmillan; 2014
- [23] Zhang Z, Zhang J, Shen HT. Rapid object detection in satellite images using convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*. 2016;**54**(11):6555-6565
- [24] Bui TD, Lofman O, Revhaug I. Machine learning applications for geomorphological mapping using GIS data and remote sensing images: A review. *Geomorphology*. 2019;**327**:91-102
- [25] Wang S, Zhang J, Li Z, Li J. Spatial prediction and accuracy assessment of land use and land cover changes using remote sensing and GIS techniques. *Remote Sensing*. 2016;**8**(11):907
- [26] Li Z, Li J, Wu F, Tian Z. A comprehensive review of geospatial big data management. *ISPRS International Journal of Geo-Information*. 2020;**9**(7):399
- [27] Church RL, Murray AT. *Business Site Selection, Location Analysis, and GIS*. Hoboken, New Jersey. USA: John Wiley & Sons, Inc.; 2009
- [28] COMPASS INFORMATICS. 2023. Available from: <https://compass.ie/services/gis-and-data-analytics/> [Accessed: August 16, 2023].
- [29] Andreev I. *Knowledge Management. VALAMIS*; 2022. Available from: <https://www.valamis.com/hub/knowledge-management>, <https://www.valamis.com/hub/institutional-knowledge>
- [30] Sui D, Goodchild MF. The convergence of GIS and social media: Challenges for GIScience. *International Journal of Geographical Information Science*. 2011;**25**(11):1737-1748
- [31] Craglia M, Annoni A, Proia F. *GSDI Cookbook: Developing Spatial Data Infrastructures*. Joint Research Centre of the European Commission. Redlands (United States of America): ESRI Press; 2008
- [32] Rajabifard A, Feeney ME. Spatially enabling government: Opportunities, barriers and future directions. *Land Use Policy*. 2013;**31**:574-583
- [33] Liang SH, Yang CF, Liu WL, Chen KC. Development and evaluation of a cloud-based geospatial information platform. *ISPRS International Journal of Geo-Information*. 2015;**4**(4):1982-2001
- [34] Janowicz K, Hitzler P, Adams B, Kolas D, Vardeman C, Henson C. On the strategic value of linking ontologies and linked data for the geospatial sciences. *Semantic Web*. 2018;**9**(4):545-563
- [35] Rinner C. *Neogeography: The challenge of augmenting the geospatial web*. *Geographic Information Sciences*. 2003;**9**(2):129-138
- [36] Dalkir K. *Knowledge Management in Theory and Practice*. London, UK: Routledge; 2011. DOI: 10.4324/9780080547367

Remote Sensing and GIS Applications in Wildfires

Georgios Zagalikis

Abstract

Wildfires are closely associated with human activities and global climate change, but they also affect human health, safety, and the eco-environment. The ability of understanding wildfire dynamics is important for managing the effects of wildfires on infrastructures and natural environments. Geospatial technologies (remote sensing and GIS) provide a means to study wildfires at multiple temporal and spatial scales using an efficient and quantitative method. This chapter presents an overview of the applications of geospatial technologies in wildfire management. Applications related to pre-fire conditions management (fire hazard mapping, fire risk mapping, fuel mapping), monitoring fire conditions (fire detection, detection of hot-spots, fire thermal parameters, etc.) and post-fire condition management (burnt area mapping, burn severity, soil erosion assessments, post-fire vegetation recovery assessments and monitoring) are discussed. Emphasis is given to the roles of multispectral sensors, lidar and evolving UAV/drone technologies in mapping, processing, combining and monitoring various environmental characteristics related to wildfires. Current and previous researches are presented, and future research trends are discussed. It is widely accepted that geospatial technologies provide a low-cost, multi-temporal means for conducting local, regional and global-scale wildfire research, and assessments.

Keywords: wildfires, GIS, remote sensing, multispectral sensors, lidar, UAV, mapping

1. Introduction

Forest fires (term used in Europe to designate the unwanted fires burning forests and wildlands) are considered as one of the main environmental hazards worldwide. Forest fires are periodic disturbance events that dramatically affect the structure and distribution of global forest ecosystems, altering soil erosion, causing loss of biodiversity, habitat, production and productivity, endangering human life, and disrupting livelihoods [1].

The impacts of climate change are becoming more evident each year. Forest fires are more frequent and extreme, as also confirmed by the recent report of the International Panel on Climate Change (IPCC) “Climate Change 2022: Impacts, adaptation and vulnerability” [2] and the recent United Nations Environment Program report “Spreading like wildfire: The rising threat of extraordinary landscape fires” [3].

In Mediterranean basin, natural forest fires are common and are an integral part of the terrestrial ecosystems (**Figure 1**). Fire has been used by man as management tool since early times. Mediterranean region is the most affected area from wildfires in Europe. Approximately 85% of the total burnt area occurs in the EU Mediterranean region [4]. This vulnerability of Mediterranean basin to be affected by fire is linked to its climate, which is characterized by rainy and mild winters followed by warm and dry summers [5]. Extreme weather conditions in summer (high temperature, strong wind, low relative humidity and drought) are a key factor in the ignition and spread of large forest fires [6–8].

Geospatial technologies such as remote sensing (RS) and geographic information systems (GIS) have been used for the management of forest fires several years ago with the use of aerial photographs and have been increasing as new type of sensors become available varying from spaceborne sensors to the newly era of UAVs sensors. Remote sensing provides the data used through the GIS or other decision support systems.

Three main levels can be distinguished, in which remote sensing and GIS provide results that can be applied directly to the subject of forest fires: pre-fire conditions (fire risk assessment, vegetation mapping, topography, risk of fire spreading, etc.), monitoring fire conditions (fire detection, detection of hotspots, fire thermal parameters, etc.) and post-fire (burnt area mapping, success of forest regeneration, flood risk, etc.). Different types of remote sensing platforms and sensors are used for data acquisition as satellites to UAVs and active sensors (SAR, Lidar) to passive sensors (optical, thermal infrared, multispectral).

The purpose of this chapter is to present a wide outline of the use of remote sensing and GIS in wildfire management. The chapter examines research in the

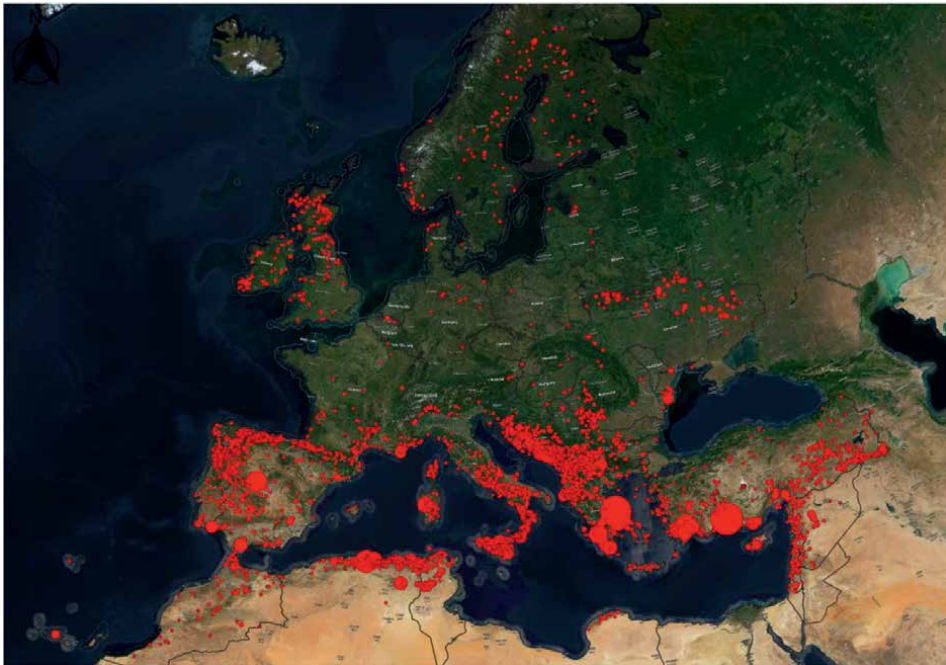


Figure 1.
Forest fires in the pan-European region in 2021 © EU, 2022.

three main levels and within each level, and sub-level, current research trends and case examples are presented, divided by the platform and sensor type used. The goal of this chapter is to provide the reader with a broad view of the current applications of remote sensing and GIS technologies and techniques in the field of wildfire management.

2. Pre-fire condition management

A forest fire results from a complex interaction of biological, meteorological, physical and social factors that influence the likelihood of a forest fire breaking out, its propagation and intensity, duration and extent, and its potential to cause damage to economies, the environment, and society [3]. Terms such as fire hazard, fire risk or danger and severity have been introduced to define risk, hazard and the characteristic (or uncharacteristic) nature of forest fires [9].

Fire risk or danger is a term used to represent the vulnerability of an area to the ignition and spread of fire. This fire vulnerability is influenced by various factors that are dynamic and static in both spatial and temporal dimensions; therefore, it is a complex concept to measure [10]. Furthermore, fire risk factors are both biotic and abiotic [11]. Biotic conditions are aimed at studying the morphological (i.e., fuel biomass load, species composition, height and density) and physiological (i.e., moisture status and chemical properties) characteristics of vegetation, while abiotic indicators are related to the study of external conditions such as topography and meteorological factors on the other hand [11, 12].

Fire risk assessments involve identifying the potentially contributing variables and integrate them into mathematical expression known as “index” [13]. Several risk indices have been developed for fire risk evaluation based on either the temporal scale [14, 15] or the variable data sources [16].

Mapping is used to visualize the fire risk assessments, and this is accomplished through two primary methods: (i) point-wise meteorological data-based operating systems and (ii) the use of remote sensing technologies and geographic information systems (GIS) [17].

The early methods, which assess fire danger, were based on meteorological factors since they have significant role in the occurrence of severe fire episodes [18, 19], including the Nesterov index for use in the former Soviet Union [20], the Forest Fire Danger Index (FFDI) for eastern Australia [21], the National Fire Danger Rating System for the USA [22] and the Canadian Forest Fire Weather Index (FWI) System [23]. These systems use meteorological data such as temperature, precipitation, humidity and wind speed to assess fire danger over extent geographic regions. However, these systems suffer from several limitations as the need for an extent network of weather stations or the need for interpolation to generate fire danger maps. Since the geoinformatic technologies as remote sensing/GIS have the ability to obtain continuous data for an area, the ability to analyze those data is recognized as an effective alternative or addition for the creation of fire risk maps [15, 24].

2.1 Fire hazard mapping

Fire hazard differs from fire danger as it does not include a meteorological component in the assessment [9]. Fire hazard or susceptibility maps are important in land use planning [25, 26], and that maps are useful as a valuable reference to reduce

vulnerability and could help to improve decision-making planning in ecological risk prevention [27]. Another potential use is in the insurance markets [28].

Several pre-fire conditions used in the fire hazard maps can be monitored using remote sensing [29]. Vegetation mapping (including fuel loads) is one of the first pre-fire conditions contributing to fire hazard models that are produced usually from high spatial resolution optical or radar images [30, 31]. The fuel maps may range from global-continental products derived from coarse-resolution sensors (AVHRR, Modis, Vegetation) to regional-local inventories, based on higher-resolution data (Landsat TM/ETM, radar or even Ikonos) [32].

Description of fuel properties is critical in all phases of fire management and fire science: prevention, fire danger estimation; suppression: fire behavior modeling; fire effect assessment: trace gas emissions and vegetation recovery after fire [33]. Several fuel characteristics are critical for fire propagation studies: crown bulk density, crown base height, canopy height, canopy closure, surface area to volume ratio, vertical and horizontal continuity, dead and live fuel loads, live woody loads and size of particles referring to vegetation geometry to the fuel type, which can be mapped, like classical vegetation mapping, from high spatial resolution optical or radar images [34].

Depending on the remote sensing data sources used different vegetation characteristics can be interoperated because of the different spatial, spectral and temporal resolution (**Table 1**).

Multispectral remote sensing provides fundamental inputs for the mapping of fire hazard [15, 35]. Land cover type plays an important role in evaluating the risk of a given area to fire events as it is related to fuel types and characteristics; usually, classification is fulfilled using vegetation indices. The simplest form of vegetation index is simply a ratio between two digital values from independent spectral bands. The Normalized Difference Vegetation Index ($NDVI = \frac{NIR - R}{NIR + R}$), the most commonly used ratio transformation for vegetation studies, is used to monitor photosynthetic activity and provides information on vegetation biomass and phenology [36–38]. Biomass is also estimated using various remote sensing data [39, 40].

Usually low or medium spatial resolution multispectral sensors such as Landsat-TM [41, 42] or MODIS-ASTER [43, 44], and hyperspectral sensors, such as Hyperion [45], AVIRIS [46] or MIVIS (Multispectral Infrared Visible Imaging Spectrometer) [47] and lately the Sentinel-2A [48].

High resolution commercial satellites, such as Quickbird, Pleiades-1A, IKONOS, GeoEye-1, WorldView-2,3, have become operational over the past decades, offering data at finer than 10-m spatial resolution have also been used for fuel mapping [44, 49]. Arroyo et al. [50] explored the utility of high resolution (4 m^2) satellite imagery for the production of fuel type maps using a different data processing approach and object-oriented classification, which allows for considering spatial context of adjacent pixels in the eventual generation of the fuel map. It was concluded that this approach produced higher accuracies than would have been generated using traditional maximum-likelihood classifiers. Such data can be used for pre-fire studies, but hurdles such as high cost per scene and the limitation common to most of optical sensors, the inability to penetrate forest canopies [51].

Beyond optical remote sensing satellites active sensors such as SAR and Lidar have been used to estimate vegetation characteristics that are critical for fuel type mapping, such as foliar biomass, tree volume, tree height and canopy closure [52, 53]. Lidar data is used in combination with other remote sensing data airborne or space-borne to estimate vegetation characteristics [49, 54].

Sensor	Spatial resolution	Temporal resolution	Advantages	Disadvantages	Info
Landsat MSS,TM, ETM+, OLI	15–30 m	16 days	Free and easily accessible	Lack of canopy penetration, low temporal resolution	https://landsat.gsfc.nasa.gov/satellites/landsat-8/
Sentinel-2	10–60 m	5 days	Free, relatively high spatial and temporal resolution, multiple near-infrared (NIR) bands	Lack of canopy penetration	https://sentinels.copernicus.eu
MODIS	250 m–1 km	1 to 2 days	Free and easily accessible, high temporal resolution, large area analysis	Lack of canopy penetration, coarse spatial resolution limits analysis of smaller areas	https://modis.gsfc.nasa.gov/about/
Pleiades Neo	0.3–1.2 m	1–3 per day	High spatial resolution	High cost	https://earth.esa.int/eogateway/missions/pleiades-neo
AVIRIS	4–20 m	Airborne	High spatial resolution, hyperspectral sensor	High cost, complicated data processing	https://aviris.jpl.nasa.gov/
AVIRIS-NG	0.3–4.0 m	Airborne	High spatial resolution, hyperspectral sensor	High cost, complicated data processing	https://avirisng.jpl.nasa.gov/aviris-ng.html
Hyperion	10–30 m	16 days	Free and easily accessible, hyperspectral sensor	Decommissioned 2017	https://www.usgs.gov/centers/eros/science/usgs-eros-archive-earth-observing-one-eo-1-hyperion
UAV	High centimeter-level	Airborne	High spatial resolution, hyperspectral sensor, lidar sensors	Cost it depends from the sensor, low endurance for areas 1–10 km ²	

Table 1.
List of sensors commonly used or with potential usage in forest fires research.

Remote sensing provides also topographic data as slope, aspect and elevation that all influence the risk of an area to fire ignition and/or spread [55]. In terms of terrain factors, altitude and slope can affect the moisture loss of vegetation, and in contrast, the aspect affects the amount of solar radiation received by vegetation, directly

affecting the vegetation drying degree. For forests, terrain differences lead to differences in wind, water balance and heat transfer between different areas, affecting the spread of fire [56].

Another pre-fire condition that remote sensing systems provide is the historical statistical data about the date and time of the fire, fire location (administrative district), fire size by land use category and fire cause [57, 58]. Historical data can be used to evaluate the accuracy of the fire hazard maps or the fire risk maps and the models used. Examples of remote-sensed databases of fire activity include the World Along Track Scanning Radiometer (ATSR) Fire Atlas [59], the MODIS and the VIIRS Active Fire Products [60] and the LSA SAF Fire Products [61].

In the last decade, unmanned aerial vehicles (UAVs) have become more affordable and one of the most important remote sensing data sources in various natural resource management applications [62–64]. UAV can use a variety of sensors, including visible light, near infrared (NIR), shortwave infrared (SWIR), thermal infrared (TIR), SAR and Lidar sensors. UAV optical sensors, including visible, NIR and SWIR, also record data as multispectral or hyperspectral bands [65, 66]. Since the sensor technology advances, increasingly smaller, lighter and cheaper sensors have become available for UAV remote sensing applications. Manfreda et al. [67] and Constantine and Rey [68] provide detailed introductions to various UAV remote sensing systems. Also, Dainelli et al. [69] provide a systematic review analysis of unmanned aerial vehicle (UAV)-based forestry research papers.

Remote sensing provides data about proximity to human settlement/infrastructure to identify areas at risk of fire occurrence. Forest fires in the Mediterranean Europe and in Mediterranean basin in general are mostly related to human activities. More than 90% of fires originate from either deliberate or involuntary causes [70]. Mapping ignition risk is an important tool for managers, helping to improve the effectiveness of fire prevention, detection and firefighting resource allocation [71, 72].

In Mediterranean basin, in California and in Australia, electric power infrastructure has ignited several of the most destructive wildfires in recent history, even if there is a gap in published research work on wildfire prediction when the root cause of the fire is linked to the power grid infrastructure—despite the fact that some of the most devastating wildfires were sparked by power grids [73]. Remote sensing (RS) and geographic information system (GIS) technologies and their derived applications are effective tools for power line surveillance and disaster prevention [74]. UAVs can be used for detailed map of electric power grid and for identifying utility systems that pose a fire risk [75, 76].

Several methods have been used widely for forest fire occurrence. Based on the literature, there are two main methods and techniques for this aim, namely, data-based (machine learning (ML)) and knowledge-based methods [77]. Numerous studies used knowledge-based methods for forest fire susceptibility mappings, such as fuzzy logic [78], the analytic hierarchy process (AHP) [79] and the analytical network process [80]. In recent years, an obvious trend in the use of machine learning algorithms is in a wide range of natural hazard assessment disciplines [81, 82]. Machine learning (ML) methods, such as logistic regression [83], artificial neural networks [84, 85] and Random Forest [86, 87], are also used for forest fire hazard mapping.

An example of fire hazard mapping is Fire Hazard Severity Zones Maps in California USA and it is produced by State of California and the Department of Forestry and Fire Protection. The Fire Hazard Severity Zone (FHSZ) maps are developed using a science-based and field-tested model that assigns a hazard score

based on the factors that influence fire likelihood and fire behavior. Many factors are considered such as fire history, existing and potential fuel (natural vegetation), predicted flame length, blowing embers, terrain, and typical fire weather for the area. There are three levels of hazard in the State Responsibility Areas: moderate, high and very high (**Figure 2**) [88].

2.2 Fire risk mapping

Fire risk mapping also includes dynamic variables such as weather and vegetation conditions. The result is a map that displays varying degrees of fire risk ranging from very low to very high. Fire risk maps in short term normally provide risk estimates that are only appropriate for a short period (days-weeks) after their creation. They use many of the same inputs as fire hazard maps but also include variables that are continuously changing such as fuel moisture content, weather conditions and vegetation conditions [9, 89]. Live fuel moisture content (LFMC) is one of the critical factors affecting fire ignition and fire propagation, and therefore is taken into account in most fire danger and fire behavior modeling systems [90]. Dead fuel moisture conditions also affects fire ignition and fire propagation and the computation is easily estimated from weather data and fuel characteristics, because dead fuel moisture is in balance with that of the surrounding atmosphere [91, 92]. Several studies conducted in a wide range of ecosystems have found a significant correlation between burned area and LFMC [93, 94]. Accurate and comprehensive spatial and temporal estimations of LFMC are essential to assess wildfire danger [95] and to develop early warning systems for the evolution of critical conditions [96]. Field-based estimations of LFMC are labor and time intensive as well as costly, and they cover small areas and are difficult to be frequent.

Remote sensing methods cover large areas and this is why remotely sensed images have been used extensively for estimation of the LFMC using various methods, which

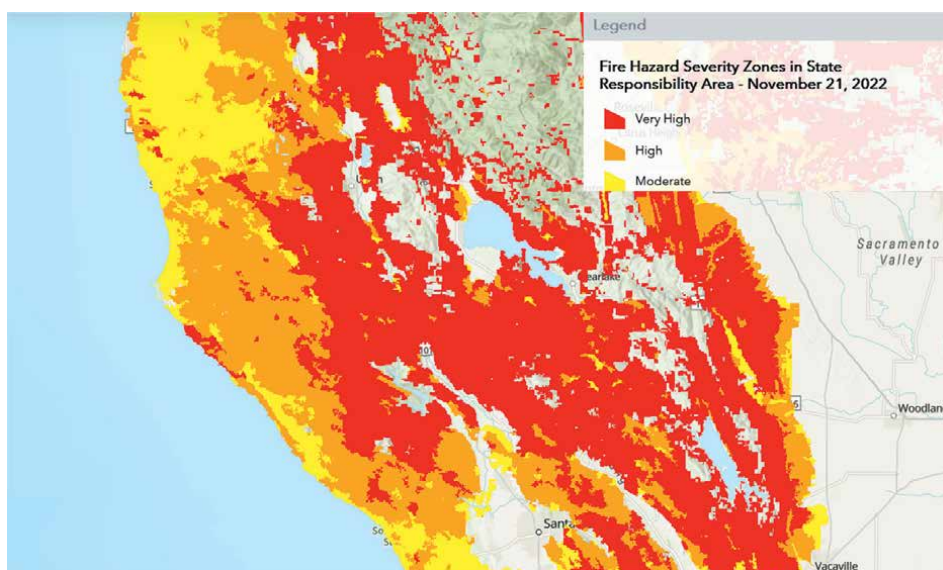


Figure 2.
California fire Hazard severity zone viewer web mapping application.

may be generally classified into statistical (empirical) [97, 98], physical model-based approaches [99, 100] and recently machine learning approaches [101, 102].

The use of satellite data in LFMC estimation has been focused on coarse spatial resolution remote sensing data such as moderate resolution Imaging Spectroradiometer (MODIS) and Advanced Very High-Resolution Radiometer (AVHRR). Combinations of spectral indices have been successfully employed to estimate LFMC [94, 97, 103] usually including the reflectance in the near infrared and the reflectance in the 1.6- μm region [104, 105]. Other studies implement also thermal data to estimate LFMC [106], different satellite data (Sentinel-2) in combination with meteorological variables or without [107, 108] and microwave remote sensing [102, 109, 110]. Yebra et al. [111] reviewed the use of remotely sensed data for estimating LFMC with particular interesting toward the operational use of LFMC products for fire risk assessment.

Fire Weather Index (FWI) module of the Canadian Forest Fire Danger Rating System (CFFDRS) is widely used in several countries such as Alaska, USA [112]; Indonesia and Malaysia [113]; Portugal [114]; Spain [115]; and others. It is also part of the European Forest Fire Information System (EFFIS) (**Figure 3**) [116]. The FWI System has six components rating fuel moisture content and potential fire behavior in a common fuel type (i.e., mature pine stand) and in no slope conditions.

Calculations are based on daily noon measurements of air temperature, relative humidity, wind speed and previous 24-h precipitation. Despite the global acceptance of the FWI, it has an inherent problem in delineating the spatial dynamics of the danger conditions, as it employs geographic information system (GIS)-based interpolation techniques [117]. Sirca et al. [118] assessing the performance of several fire danger indexes in the Mediterranean area concluded that FWI yielded better predictions than the other indexes.

In Italy, the Civil Protection adopted the RISICO model, which is a fire danger rating system that was developed specifically for the vegetation cover of the Mediterranean [119]. RISICO integrates meteorological observations and forecasts from an NWP (Numerical Weather Prediction) Limited Area Model (LAM) and ECMWF-Integrated Forecasting System (IFS) with vegetation cover and topography data. Another project developed in Sardinia, Italy called Daily Fire Hazard Index (DFHI) [120], the index is computed using satellite images and the meteorological data.

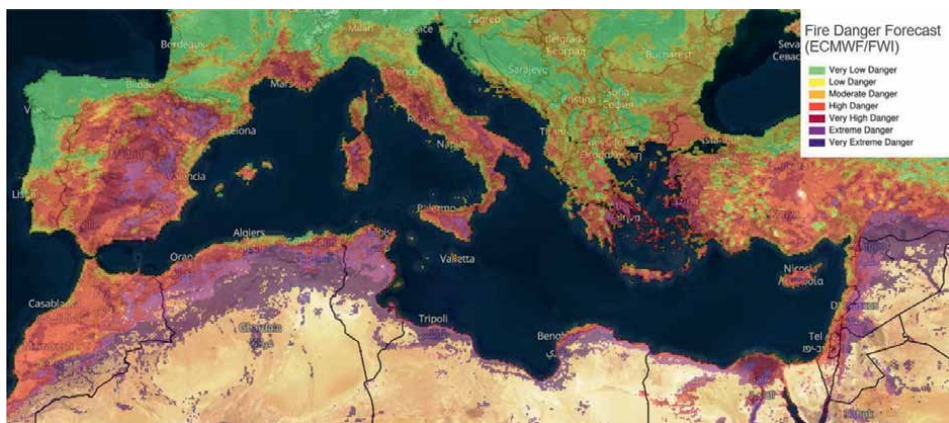


Figure 3.
Forest fire danger forecast in the pan-European region for 02 Jun 2022 © EU, 2022.

3. Fire detection and monitoring fire conditions

Fire detection is the earliest critical stage of forest fires management, which is aimed at either fighting or monitoring the fire. For firefighting the early detection is essential; so far, fire detection is based on human observation, the use of fixed optical cameras to monitor the surrounding environment or aerial survey [29]. The temporal and spatial analysis capabilities provided by current satellite sensors are not considered sufficient for firefighting operations by forest fire managers.

3.1 Spaceborne multispectral sensors

Several of geo-stationary satellite sensors have been used for active fire detection and monitoring including but not only: Advanced Very High-Resolution Radiometer (AVHRR), Meteosat Second Generation, Spinning Enhanced Visible and Infra-red Imager (MSG-SEVIRI), Himawari-8, Geostationary Operational Environmental Satellite (GOES), DMSP OLS (US Air Force Defence Meteorological Satellite Programme – Operational Line-scan System), Visible Infrared Imaging Radiometer Suite (VIIRS) and moderate resolution imaging spectroradiometer (MODIS) [121–123]. Early attempts of active fire detection and monitoring started during 80' with the US Forest Service using NOAA satellites to identify forest fires in the western US [124]. AVHRR was the primary sensor system for active fire detection and monitoring until the eventual launch during 1999 of the MODIS sensors onboard the Terra and Aqua platforms [125]. MODIS Terra and Aqua have become the primary sensors at regional to global scales for active fire detection and monitoring due to their high temporal resolution and special channels designed for fire monitoring. The Aqua MODIS instrument acquires data twice daily (1:30 PM and AM), as does the Terra MODIS (10:30 AM and PM). These four daily MODIS fire observations serve to advance global monitoring of fire processes and their effects on ecosystems, the atmosphere and climate [126]. Several operational fire monitoring systems using MODIS active fire detection include the Canadian Wildland Fire Information System (CWFIS) (<http://cwfis.cfs.nrcan.gc.ca>), the USA Active Fire Mapping Service or the European Forest Fire Information System (EFFIS) (<http://effis.jrc.ec.europa.eu>). In the case of EFFIS, post-processing filters based on landcover ancillary data are applied to the MODIS product to reduce the number of false alarms produced by non-fire hot surfaces (e.g., industrial areas, hot ground soils) and therefore increase the reliability of the active fire detection [116]. **Figure 4** shows active fires monitoring in the wider Mediterranean region.

Several studies found that MODIS fire detection algorithms had trouble detecting smaller/cooler fires and frequently detected false alarms [127, 128]. Validation of MODIS and AVHRR active fire products used higher spatial resolution sensors such as Landsat TM and ASTER to evaluate the performance of the MODIS and AVHRR products [129, 130]. The accuracy of fire detection is calculated in terms of commission and omission errors as well as error distributions over various locations. Commission and omission errors have been reported by many studies errors of omission represent “failure to detect fires” and errors of commission represent “false alarms or false positives” pixels identified as fires [130]. Schroeder et al. [127] found errors of commission of approximately 35% over areas of active deforestation, and similarly, Forghani et al. [130] report that overall commission errors of MODIS and AVHRR hotspots over the 5% sample data were 15 and 68%, respectively, and overall omission errors of MODIS and AVHRR hotspots were 17 and 23%, respectively. An interesting comparison of the



Figure 4.
Active fires from MODIS during June 2022 © EU, 2022.

suitability of some of geo-stationary satellite sensors (AVHRR, MODIS, DMSPOLS) for fire detection was carried out by Cahoon et al. [131]. They found in their study that the minimum size for a detectable fire was approximately 213 m² for MODIS, 435 m² for AVHRR and 45 m² for DMSP-OLS, for a nominal fire front temperature of 1000 K and a background temperature of 305 K. Maier et al. [132] determined a minimum fire size of 100–300 m² for the detection by the MODIS algorithm. Evaluation of the MODIS active fire product to quantify detection rates of both Terra and Aqua MODIS sensors was carried out by Hawbaker et al. [133]. Finally, MODIS had the ability to capture large fires in the US, but may under represent fires in areas with high cloud, rapid burning or small- and low-intensity fires that are often undetected.

To address these limitations found during evaluation of the MODIS fire detection algorithm, NASA's Earth Observing System periodically reprocesses the raw instrument data archive using newer fire detection algorithms. In the MODIS version 3 active fire detection algorithm's sensitivity to small fires was sacrificed to reduce false alarms over certain surface types during the day [134]. The MODIS 4, contextual fire detection algorithm was enhanced, which increased the ability to detect small cool fires [60]. For the recently released Collection 6 MODIS fire products, bands 1 (red), 2 (near infrared), 7 (short-wave infrared), 21 (thermal infrared over a wide-range centered on 4 μ m), 22 (thermal infrared in a narrow-range centered on 4 μ m), 31 (thermal infrared at 11 μ m) and 32 (thermal infrared at 12 μ m) are used to identify 1-km "fire pixels" that contain at least one active fire at time of acquisition [128].

Further than MODIS, recent studies have used data from the Landsat 8 Operational Land Imager (OLI) [135] and Sentinel-2 sensors [136]. Also, the use of geostationary weather satellites and their ability to detect active fires have been a research topic of great interest for several years [137–139].

Active fires can be detected by a wide variety of spectral sensors, but as active fire products are usually used to monitor ongoing fires and the ideal sensors for detecting them will possess a high temporal resolution; at the same time that kind of sensor have low spatial resolution. While sensors with higher spatial resolution like Landsat satellites and Sentinel-2 satellites work well for detecting active fires, their low temporal resolution makes them less than ideal. This is the reason why most active fire products are generated from sensors such as MODIS and VIIRS, as these sensors have

a very short revisit time (<2 days) allowing for the active fire products to be rapidly updated [17]. Geostationary satellites have even higher temporal resolution, with updates every few hours or less. However, these high temporal resolutions come at the cost of lower accuracies due to the omission of smaller/cooler fires and commission errors caused by activities such as deforestation [127, 132, 140].

3.2 UAVs

The MODIS satellite imagery with the high temporal resolution (1–2 days revisiting time) is commonly applied in forest fires detection and monitoring. However, the low spatial resolution of MODIS is insufficient for this task at local scales [65]. During the last decade, we have seen a great progress in the field of using unmanned aerial vehicles (UAVs) for forest fire monitoring, detection and even fighting [141]. UAVs have become smaller, more affordable and now have better computation capabilities than in the past making them reliable tools for remote sensing missions in hostile environments [142]. UAVs provide rapid, mobile and cost-effective sensor systems that are currently being adopted for fire detection and monitoring [143, 144]. In their review of UAVs applications for forest fire monitoring, detection and fighting, Yuan et al. [144] define fire monitoring as involving the active search for possible fire occurrences, while fire detection involves the identification of fires in progress. UAVs used for fire monitoring and detection rely on sensors operating in the visible and thermal infrared wavelengths. In a more recent review of UAVs usage in forest fires, Akhloufi et al. [145] present a survey of different approaches for the development of UAV fire assistance systems. They conclude that UAVs can play an important role in the fight against wildland fires in large areas and with the decrease in their prices and their wider commercial availability, new applications in this field will emerge. However, limitations remain such as autonomy, reliability and fault tolerance; also security issues are a concern, as there are risks associated with having UAVs flying over firefighters or close to aircraft carrying water and fire retardants [145].

4. Post-fire condition management

The application of remotely sensed imagery to monitor and assess the impacts of fire on local and regional environments can be broadly divided into several applications such as burnt area mapping, burn severity assessments, vegetation recovery monitoring, effects on soil erosion, and general fire effects on air quality, soil, vegetation and fauna.

4.1 Burnt area mapping

Remotely sensed data have been extensively used for burnt area mapping at local, regional and global scales [146, 147]. Burned area mapping is critical importance for forest managers, climate scientist and policy makers; they provide accurate spatial representations of fire extents and perimeters. Accurate maps of the areas affected by forest fire are needed for rehabilitation planning, calculating the economic and environmental cost of fires, and for regional and global scale estimates in gas and particulate emissions [148, 149].

Fires produce a significant change in the structure and the reflectance of vegetation and the soil properties within the burnt area that are noticeable in the microwave,

visible and especially the infrared part of the electromagnetic spectrum [29]. The sensors used for burned area mapping differ depending on scale and purpose of the assessment. Since the 1980s, the majority of techniques have been developed for data acquired from the AVHRR sensor, and as such were restricted to a limited number of reflectance and thermal bands; most commonly method used is multi-temporal comparison of NDVI [150, 151]. Although data from the AVHRR sensor have low spatial resolution (i.e., 1.1 km), global data have been obtained from a series of different satellites since the 1980s allowed the long-term monitoring of large-scale fires in remote and isolated areas. During 2000s, other global burnt area datasets were derived from SPOT Vegetation and the ATSR-2 on board of Envisat [152–154]. Partial validations of the global burnt area products were performed by Roy et al. [155], Boschetti et al. [156] and Roy and Boschetti [157]. Pereira et al. [150] showed that the accuracy of the results for mapping burnt areas with AVHRR data in the Mediterranean region of Europe was about 80% for large fires. The methods were considered suitable only for fires larger than 1000 ha and reliable for fires larger than 2000 ha. However, the mapping of those fires would correspond only to approximately 30 and 21%, respectively, of the total yearly burnt area in the European Mediterranean region.

Until the launch of the MODIS sensor on board of the TERRA and AQUA satellites, there was not a space-based system design specifically to “look” at terrestrial Earth. MODIS sensors add a new capability for regional mapping of burnt areas, the availability of free data of medium spatial resolution from the MODIS sensors since 2000 provided a definite impulse for the use of remote sensing at the regional and global scales [134]. The higher spectral information and the better radiometry of the MODIS sensor in comparison with the previous sensor systems (e.g., AVHRR—an atmospheric mission), provided the right data for the mapping of burnt areas at these scales. At the global scale, the MODIS program has used to produce a standard product on burned areas based on a multi-temporal change detection approach to analyze differences between modeled and actual reflectance, using an algorithm based on Bidirectional Reflectance Distribution Function (BRDF) change detection approach [155]. Giglio et al. [147] created an algorithm for burned area mapping using 1 km MODIS data, the normalized burn ratio (NBR) and temporal texture. The resulting algorithm (MCD64) was assessed based on Landsat-derived burned area maps for central Siberia, the western US and southern Africa. They found that the algorithm performed well overall, except in a closed canopy subregion in southern Africa where it underestimated burned area. In general, global burned area products have been shown to exhibit relatively large errors of omission and commission. The accuracies of several of these burned area products were compared by Padilla et al. [158]. Six global burned area products were compared using stratified random sampling in the first attempt to implement a statistically designed sample to validate burned area products on a global scale. The products used in this study included MCD45, MCD64, Geoland2, MERGED_cci, MERIS_cci and VGT_cci. The study found MCD64 to be the most accurate, followed by MCD45; however, all the products possessed burned area commission errors above 40% and omission errors above 65%.

At regional scale, MODIS is operationally used in systems such as Canadian CWFIS and the European EFFIS. Two full mosaics of MODIS data are received and processed daily in EFFIS to provide near-real time monitoring of wildfires and map burnt areas. The systems are thus updated up to two times daily, providing accurate information of fire impacts in Europe [159]. Rapid Damage Assessment (RDA) module of EFFIS was initially implemented in 2003 to map burned areas during the fire season, by analyzing MODIS daily images at 250-m spatial resolution. Since the

year 2016, the RDA incorporates the mapping of active fires and burnt areas from the VIIRS Sensor, onboard the NASA Suomi National Polar-orbiting Partnership (SNPP) and the NOAA-20, which allows the update of burnt areas maps one more time, every day. The EFFIS from 2018 is started using also Sentinel-2 imagery at 20-m resolution, which allows the detection of fires below the 30-ha threshold and it is estimated that the areas mapped in EFFIS represent about 95% of the total area that burns in the EU every year. **Figure 5** shows the extent of burnt areas as they were mapped from MODIS and Sentinel-2 imagery during the last fire season and **Figure 6** shows details about the area burnt (dates, location, area, type of vegetation affected).

At national to local scales, the wide variety of remotely sensed products at high to moderate resolution (1-m to 30-m ground spatial resolution) makes it possible the accurate mapping of burnt areas. Optical image satellites such as the Landsat-8 and Sentinel-2 constitute the latest generation Earth observation missions of the United States Geological Survey and European Space Agency, respectively. The Landsat-8 satellite, launched on February 11, 2013, has 11 bands ranging in spatial resolution from 15 m (panchromatic) to 30 m (visible, near infrared, short-wave infrared) and then to 100 m (thermal). Within 16 days' temporal resolution, Landsat-8 provides 30-m spatial resolution optical imagery on eight spectral bands *via* the Operational Land Imager sensor, which can be accessed freely [160, 161]. Landsat-8 30-m products can map small and spatially fragmented burned areas greater than 40 ha in size with its more detail than other satellite sensors, such as the MODIS 500 m [162]. Sentinel 2 is a high-resolution multispectral imaging satellite composed of Sentinel-2A and Sentinel-2B satellites. The Sentinel-2 satellite covers 13 spectral bands with ground resolutions of 10 m, 20 m and 60 m and provides red-edge spectral bands, the satellite temporal resolution 10 days [163]. Filipponi [164] using threshold-based classification for burnt area mapping of on the 2017 Italy forest fires based on Sentinel-2 time series data reported the methodology generated commission error of around 25% and an omission error of around 40%.

Since the Sentinel-2 and Landsat-8 have similar wavelengths and the same geographic coordinate system, several studies combined both sensors for burnt area mapping [162, 165, 166]. Syifa et al. [166] using SVM and SVM-ICA algorithms for burned area mapping for the Camp Fire (California, USA) report that the SVM-ICA



Figure 5.
Burnt area mapping MODIS/SENTINEL2 from the last fire season © EU, 2022.

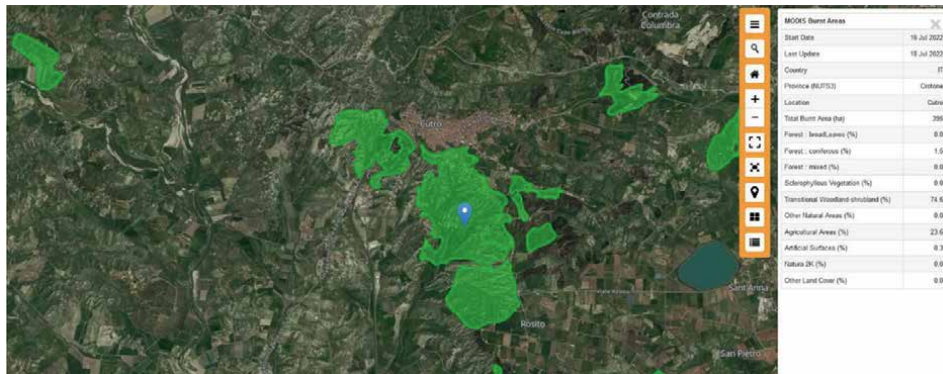


Figure 6.
Burnt area mapping from a fire occurred at July 16, 2022 near Cutro, Crotone, Italy © EU, 2022.

produced a better accuracy (overall accuracies of 83.8 and 83.6% for pre- and post-wildfire using Landsat-8, respectively; 90.8 and 91.8% for pre- and post-wildfire using Sentinel-2, respectively), compared to SVM without optimization (overall accuracies of 80.0 and 78.9% for pre- and post-wildfire using Landsat-8, respectively; 83.3 and 84.8% for pre- and post-wildfire using Sentinel-2, respectively). Different methods of satellite-based burned area detection have been developed, including threshold-based methods using multi-spectral bands or spectral indices such as Normalized Difference Vegetation Index (NDVI), the Soil-adjusted Vegetation Index (SAVI), and the Normalized Burned Ratio (NBR), supervised classification, logistic regression, random forest, etc. [162, 165, 167]. Reviews of burned area mapping algorithms are found in the studies by Boschetti et al. [168] and Chuvieco et al. [169]. Synthetic Aperture Radar (SAR) was also used for burned area mapping mainly in boreal or tropical regions [170, 171], but some examples for the Mediterranean area exist [172, 173]. Also, SAR data have been used in combination with optical data for burned area mapping [174].

4.2 Burn severity and soil erosion assessments

The term fire or burn severity was born out of the need to provide a description of how fire intensity affected ecosystems, particularly following forest fires where direct information on fire intensity was absent and effects are often quite variable within and between different ecosystems [175]. Burn severity impacts vegetation mortality and soil nutrient composition, and causes increased run-off due to decreased of organic matter in the soil and the aboveground vegetation [176, 177]. Burn severity is usually measured in the field using the composite burn index (CBI), which involves an optical assessment of burned areas to determine the fire impacts on ecological conditions. Due to the need for a systematic approach to estimate burn severity across different environments, this approach uses a measure called the composite burn index (CBI) designed to provide a single index to represent many different phenomena of interest to land managers [178]. The CBI combines fire severity metrics and ecosystem responses that include resprouting of herbs, shrubs and hardwood trees, and seedling colonization. The CBI was created to allow for visual estimates to be conducted by rating the degree of damage done by the fire, as well as the estimated vegetation recovery for the area, on a 0 to 3 scale [179].

Various remote sensors (e.g., MODIS, AVIRIS) have been tested for their ability to match field measurements of severity and the Landsat Thematic Mapper sensor was widely accepted as most appropriate for this task [180–182].

Spectral indices, such as the Normalized Burn Ratio (NBR) and NDVI, have been widely used for assessing burn severity through remote sensing in boreal regions [183–186]. The NBR has mostly replaced the NDVI as the standard index for burn severity assessment and been widely used as a means for approximating the burn severity and burned area using satellite imagery also in the Mediterranean basin [187, 188]. The NBR is calculated using NIR and SWIR data, with the SWIR wavelength interval generally within the 2.08–2.35- μm range [189]. A numerous differenced indices for burn severity assessment using multi-temporal satellite data and bi-temporal image differencing techniques have been used in boreal regions, such as the Differenced Normalized Burn Ratio (dNBR) [186, 190], the Differenced Normalized Difference Vegetation Index (dNDVI) [191, 192] and a relative version of the dNBR (RdNBR) [193]. The majority of this indices used NIR and SWIR data, where the NIR wavelengths are sensitive to the leaf structure of live vegetation, while the SWIR is sensitive to moisture content and some soil conditions [189, 194]. Vegetation affected by a fire show decreased NIR reflectance and increased SWIR reflectance [187].

Temporal differences in land surface albedo (LSA) and land surface temperature (LST) have been used to estimate burn severity [195, 196]. Quintano et al. [195] found that changes in LST showed high agreement with field-measured CBI when used to map burn severity for an ecosystem dominated by maritime pine (*Pinus pinaster*) in Sierra del Teleno, Spain. Zheng et al. [196] conclude that the results indicated that the $\Delta\text{LST}/\text{EVI}$ performed relatively better than some commonly used burn severity indices, and its performance was stable and steady across regional forest areas of different eco-type after fires. However, Veraverbeke et al. [197] and Quintano et al. [195] found that while changes in LSA and LST were highly related to burn severity they were highly dependent on seasonality.

With the launch of the Sentinel-2 sensors by the European Space Agency (ESA), with specific spectral bands to record data in the vegetation red-edge spectral domain, which is one of the best radiance-based descriptors of chlorophyll content [198], allowed for the development and application of new spectral indices to discriminate burn severity.

Fernández-Manso et al. [199] used Sentinel-2A imagery to examine capabilities of red-edge spectral indices in the context of burn severity in central-western Spain. The area is a mixed of shrubs and trees, with maritime pine (*Pinus pinaster*) and Pyrenean oak (*Quercus pyrenaica*) being the dominate species. Results indicated that the red-edge spectral indices calculated using red-edge 1(B5) and red-edge 3(B7)/NIR(B8) were most suitable for burn severity assessment. Filipponi [164, 200] tested Burned Area Index for Sentinel-2 (BAIS2, BAIS2 α), based on Sentinel-2 spectral bands on various study cases in Italy for summer 2017 fires, and results show a good performance of the index and highlighted critical issues related to the Sentinel-2 data processing. Burn severity assessments require imagery with a high to moderate (<100 m) spatial resolution such as Landsat series and the Sentinel-2 sensors, and the limitation of these sensors is the long temporal resolution, which can make the rapid acquisition of data on post-fire conditions difficult.

With the use of more advanced remote sensing technologies, such as hyperspectral imagery, lidar, radar and UAVs, important fire effects may be more accurately and consistently inferred from imagery with higher spectral, spatial and temporal resolutions [201]. Lidar sensors provide a relative new technology, which can be used

for burn severity assessments. Lidar data can provide pre- and post-fire vegetation structure profiles, which can be used for severity of burning estimations. Wang and Glenn [202] showed the potential for lidar-derived burn severity estimates in sagebrush steppe rangelands using average vegetation height change. This method outperformed the dNBR index estimations, with an overall accuracy of 84%, and proved to be sensitive to differences between moderate and high severity burns. Wulder et al. [203] compared changes in boreal forest structure, obtained by lidar returns, to post-fire conditions, estimated using spectral indices for the Boreal Plains ecozone in Alberta, Canada. The researchers found that absolute and relative changes in post-fire forest structure exhibited a high correlation with post-fire conditions. The synergy of lidar with multispectral sensor data for burn severity assessment is a topic of ongoing research [204–206]. Viedma et al. [206] crossed high-resolution lidar data, acquired from an UAV after the fire and fire severity levels based on the relativized burnt ratio (RBR) derived from Sentinel 2A images acquired a few months before and after fire. They conclude that lidar metrics derived from vertical canopy profiles (VCPs) demonstrated promising potential for characterizing fine-grained post-fire plant structures and fire damage when crossed with satellite-based fire severity metrics, turning into a promising approach for better characterizing fire impacts at a resolution needed for many ecological processes.

UAV technologies have seen an important progression in the last decade and they are now used in a wide range of applications. This technology has recently been applied to research focusing on burn severity assessment using various sensors [206–213]. Ye et al. [211] used object-oriented method to determine the optimal segmentation scale for forest burn severity and a multilevel rule classification and extraction model is established to achieve the automatic identification and mapping. The imagery was obtained by a small, multi-rotor near-ground UAV and the results show that the method mentioned above can recognize different types of forest burn severity: unburned, damaged, dead and burnt with overall accuracy been 87.76%. Carvajal-Ramírez et al. [207] employed an UAV carrying a high-resolution multispectral sensor including green, red, near-infrared and red-edge bands. Flights were carried out pre- and post-controlled fire in a Mediterranean forest. The products obtained from the UAV were a Digital Surface Model (DSM) and multispectral images orthorectified in both periods, temporal differences (d) between pre- and post-fire values of the Excess Green Index (EGI), Normalized Difference Vegetation Index (NDVI) and Normalized Difference Red Edge (NDRE) index, where used to after reclassification to produce fire severity classes. It was concluded that dNDVI was the index that best estimated the fire severity according to the UAV flight conditions and sensor specifications.

Forest fires can negatively impact water catchments, contaminating, increasing soil erosion, changing soil composition and slope stability for extensive periods. A major reason for post-fire assessments of fire or burn severity is that it is used as an important indicator of the potential water runoff and erosion [214–216], the loss of vegetation exposes soil to erosion and makes burned areas more vulnerable to runoff and then susceptible to flood [217].

Several studies used fire or burn severity data derived from remote sensing applications or/and GIS applications for post-fire soil erosion assessment [217–219]. Lanorte et al. [218] used a distributed model based on the Revised Universal Soil Loss Equation (RUSLE) to estimate potential post-fire soil loss for four different fire events occurred in Basilicata region (Southern Italy) during 2017. GIS techniques and remote sensing data have been adopted to build a prediction model of post-fire

soil erosion risk. Results show that this model is not only able to quantify post-fire soil loss but also to identify the complexity of the relationships between fire severity and all the factors that influence soil susceptibility to erosion. Fox et al. [217] used POSTFIRE a GIS-based model that maps the burn scar and quantifies fire impacts on runoff and soil erosion. The model used pre-fire and post-fire satellite images, calculates the impact of a fire on total rainfall event runoff, and maps soil erosion rates. Additionally, the model requires a Digital Elevation Model (DEM), a mask of the general contour of the fire, land cover map, and tables of runoff coefficients and sediment concentration values for the land covers. Preliminary results from two fires in SE France indicate the model is simple to use, adaptable to local conditions, and provides realistic outputs for the burn scar, discharge and soil erosion maps after a forest fire. UAV imagery has also been used for soil erosion modeling in various types of landscape [220–222].

4.3 Post-fire vegetation recovery monitoring

Vegetation recovery after a fire event is an important measurement for the determination of the long-term impacts that fire has on an ecosystem. During the process of recovery, vegetation may also be influenced by several environmental factors, such as fire severity/damage and climatological extreme events. Vegetation recovery depends on several climate factors, such as the occurrence of droughts, which may inhibit vegetation growth, but also on precipitation of high intensity, which contributes to nutrient loss and erosion by runoff [223]. Anthropogenic factors such as grazing can also affect vegetation recovery [224, 225].

Monitoring post-fire vegetation recovery is crucial, as it provides valuable information for analyzing ecosystem resilience, for determining landscape dynamics and for forest management purposes. Post-fire monitoring of vegetation recovery can be conducted on the field using plots sampling where they measure seedling germination, plant survival and restoration, and vegetation characteristics [226–228], or with the use of remote sensing technologies. Since field measurements of post-fire vegetation recovery are a very challenging and expensive task, remote sensing techniques are a time- and cost-effective way to monitor post-fire ecosystem recovery [229]. Remote sensing techniques for estimating vegetation recovery can be grouped into three categories: (1) image classification, (2) vegetation indices (VIs) and (3) spectral mixture analysis (SMA) [230].

The most practical way to monitor changes over large areas and periods is through image-processing techniques based on change detection or classification techniques [231]. Stueve et al. [232] used supervised classification to identify patterns of alpine tree recovery in Mount Rainer National Park, Washington, USA. They performed classification analysis of 1970 satellite imagery and 2003 aerial photography to delineate establishment. Local site conditions were calculated from a lidar-based DEM, ancillary climate data and 1970 tree locations in a GIS. The supervised classification method proved successful, which can be credited to the very high spatial resolution of the data used in this study. Salvia et al. [233] used a combination of unsupervised classification and field data to successfully examine the influence of burn severity on vegetation cover and soil property recovery in the wetlands of the Paraná River Delta, Argentina.

Geographic object-based image analysis (GEOBIA) is an alternative method for image classification and uses geographic objects instead of pixels as the spatial unit of analysis [234]. GEOBIA are techniques that use both spectral response and contextual

information to assess post-fire vegetation characteristics in groups of pixels (geographic objects generated by image segmentation). Polychronaki et al. [235] performed GEOBIA-based classification on multi-temporal Système Pour l'Observation de la Terre (SPOT) and European Remote-Sensing (ERS) (C-band VV) images covering the time period from 1993 to 2007 of Thasos, Greece, to estimate forest regeneration after two different forest fires occurred 19 and 23 years ago. Using data from field-stratified sampling, they found that the classification results were validated, achieving an overall accuracy of 90.5%. In the same area, Mitri and Gitas [236] used GEOBIA on a combination of very high spatial (VHS) resolution (QuickBird) and hyperspectral (EO-1 Hyperion) imagery to investigate *Pinus brutia* and *Pinus nigra* regeneration. After classification of the segmented imagery, validation showed an overall accuracy of 83.7%.

In relation to post-fire vegetation recovery, these indicators used generally rely on greenness measurements of red-near-infrared (R-NIR) vegetation indices [237, 238]. Several spectral indices based on the NIR is used for post-fire vegetation recovery estimations; nevertheless, the Normalized Difference Vegetation Index (NDVI), the Enhanced Vegetation Index (EVI) and the Normalized Burn Ratio (NBR) are the ones used most frequently [239]. Veraverbeke et al. [240] tested the utility of 13 red-near infrared (R-NIR) vegetation indices (VIs) spectral indices to detect and estimate vegetation recovery. The study found that the soil-adjusted vegetation index (SAVI) outperformed the NDVI in areas with a single type of vegetation, NDVI outperformed the SAVI in areas with heterogeneous vegetation cover and a single soil type, and overall, the NDVI was the most robust VI for assessing vegetation recovery. Several studies of post-fire vegetation responses are based on the discrimination of spectral bands and vegetation indices (mostly NDVI, dNBR and EVI) by MODIS, Landsat, SPOT and Sentinel multi-temporal imagery in different regions and forest ecosystems of the world [241–245].

Spectral mixture analysis (SMA) have also used in vegetation recovery monitoring [246, 247]. SMA studies have shown high correlation between mapped vegetation recovery and field sampling [248, 249]. There is disagreement on whether the SMA or spectral index-based approach is most effective at estimating vegetation recovery [17]. Specific nature of the vegetation's spectral response and the different components of forest recovery, the use of a single method may not be the best option, requiring multiple methods to be deployed in each ecosystem. Riaño et al. [248] and Vila and Barbosa [250] compared Spectral Mixture Analysis (SMA) against quantitative vegetation indices (NDVI and Modified Soil Adjusted Vegetation Index (MSAV)), in different types of ecosystems, such as chaparral shrub communities SMA performed better in the contrary NDVI was most accurate applicable to coniferous forests with maquis.

Further than passive remote sensing active synthetic aperture radar (SAR) has been used for post-fire vegetation recovery assessments [30, 251–254]. Laurin et al. [30] used Cosmo-SkyMed data in a Mediterranean protected area covered by maquis to detect the burnt area extension and to conduct a mid-term assessment of vegetation regrowth. The positive results obtained in this research highlight the importance of the very high-resolution continuous acquisitions and the multi-polarization information provided by COSMO-SkyMed for monitoring fire impacts on vegetation. Several studies also confirmed that the combination of SAR and optical information was essential for estimation of regeneration stages in different forest ecosystems [255–257] and for post-fire vegetation regrowth [258, 259]. De Luca et al. [259] integrating the use of Synthetic Aperture Satellite Radar (SAR) (Sentinel-1) and optical (Sentinel-2)

image time series in a Mediterranean ecosystem. They conclude that the use of optical short-wave infrared (SWIR) and SAR C-band-based data revealed that some ecological characteristics, such as the woody biomass and structure, recovered at slower rates, comparing to those suggested by using near-infrared (NIR) and red-edge data and they proposed the optimized burn recovery ratio (BRR) for the estimation and mapping the spatial distribution of the degree of vegetation recovery.

Lidar sensors are capable of penetrating through vegetation and recording forest structural characteristics [260] makes the system ideal for monitoring vegetation recovery. Sato et al. [261] estimated post-fire changes in forest canopy height and biomass using airborne Lidar in western Amazonia. Comparing burned and unburned as control sites found that even 10 years after the occurrence of an understory fire event, burned forests had significantly lower biomass and height than control sites. Gordon et al. [262] used lidar data to measure post-fire mid-story vegetation regrowth of open forests in Australia using also field surveys. They found that the metrics computed with the lidar data had moderate to strong positive associations with field-derived metrics and provided a suitable representation of post-fire vegetation cover. Several studies also have examined the combined use of optical sensor imagery and lidar data [205, 263–266]. Viana-Soto et al. [266] combine lidar with Landsat imagery to extrapolate forest structure variables over a 30-year period (1990–2020) to provide insights on how forest structure has recovered after fire in Mediterranean pine forests. They found that less than 50% of the area completely recovered to the pre-fire structure within 26 years, and the area subjected to fire recurrence showed signs of greater difficulty in initiating the recovery. They conclude that the use of Landsat data provides a unique opportunity to analyze the evolution through decades, but the 30-m pixel size may conceal a larger structural variability within. Even so, cover and height are important indicators of forest recovery that can be derived to larger areas, providing useful information to support post-fire restoration activities [267].

In addition to satellite or the conventional airborne sensors, the value of using unmanned aerial vehicles (UAV) for estimation and monitoring of post-fire forest recovery has also been incised with the use of a variety of sensors [206, 210, 268–274]. Pádua et al. [268] evaluated effectiveness for Post-Fire Monitoring of Sentinel-2 time series data in comparison with high-resolution UAV-based data in an area affected by a fire in north-eastern Portugal. Sentinel-2 images with 10-m resolution, different spatial resolutions of the UAV-based data (0.25, 5 and 10 m) were used and compared to determine their similarities. The results demonstrated the effectiveness of satellite data for post-fire monitoring, even at a local scale as more cost-effective than UAV data. Talucci et al. [273] evaluate the ability of two vegetation indices derived from UAV imagery, one based on the visible spectrum (GCC; Green Chromatic Coordinate) and one using multispectral data (NDVI; Normalized Difference Vegetation Index), to predict field-based vegetation measures collected across post-fire landscapes of high-latitude Cajander larch forests. Findings show the utility of UAV data for NDVI in this region as a tool for quantifying and monitoring the post-fire vegetation dynamics in Cajander larch forests.

UAV systems can also be used to deliver temporal digital surface models (DSM) to detect post-fire changes in vegetation recovery, and the DSMs are produced either from lidar or using photogrammetric approach from digital imagery. Qi et al. [210] used drone laser scanning (DLS) and mobile laser scanning (MLS) to describe post-fire forest structures. Their results demonstrate that fused DLS-MLS point clouds can be effective in quantifying post-fire tree structures, which facilitates foresters to develop site-specific management plans. Aicardi et al. [272] used change detection

analysis through a time sequence (2008–2015) of DSMs obtained from lidar and digital images (UAV), for a forest stand composed primarily of Scots pine (*Pinus sylvestris*) in the Aosta Valley Region, Italy. They conclude that these preliminary results highlight the usefulness of low cost, highly flexible UAV systems within areas affected by natural or anthropogenic disturbances, given the possibility to fly on-demand (e.g., immediately after the event) and quickly repeat measures when requested for environmental monitoring.

5. Discussion and conclusions

This chapter aimed to provide a wide review of remote sensing and GIS applications in the field of forest fires—wildfires. The topics include fire hazard and fire risk mapping, fuel mapping, active fire detection and monitoring, burned area estimations, burn severity assessments and mapping, soil erosion assessments and post-fire vegetation recovery assessments and monitoring. Fire hazard mapping can be improved with better mapping and monitoring of manmade Infrastructure like power lines using UAVs with active (lidar) or passive sensors. For fire risk mapping alternative geocomputational techniques as neural networks, classification and regression trees (CARTs), fuzzy modeling, and other machine learning techniques are currently a research subject [275]. Locally fuel mapping can be delivered from UAVs with active (lidar) or passive sensors in more detailed resolution than satellite data. Active fire detection and monitoring is limited from the low temporal or spatial resolution of the available satellite sensors. Smaller and cooler fires have proven difficult to detect in global active fire datasets and by geostationary sensors [132, 140]. Probably, in the near future, geostationary satellites with higher spatial resolutions may become available, allowing for near real-time detection and monitoring of small/cool fires, already new generation satellites for forest and grassland fire detection have been launched such as the Himawari-8 launched by Japan, GOES-R/S/T by the United States and Geo-Kompsat-2A (GK-2A) by South Korea, compared with the previous satellites, and these new-generation geostationary meteorological satellites have been improved in spatial resolution and observation frequency [276]. UAVs equipped with infrared or thermal or RGB cameras can be used where data can provide live feeds and be used to predict information such as propagation of a fire [277]. For burned area estimates mostly Landsat, MODIS, Sentinel, AVHRR and SPOT satellites are used, with the NDVI and dNBR the most frequently used vegetation indices compared to other indices. Relatively newer sensors as Sentinel-2 have improved the high rates of omission errors, common when using other satellite sensors, and allow the detection of small fires (< 100 ha) accounting for a significant proportion of total burned area globally [278]. Currently, burn severity assessments and mapping are mostly based using spectral indices such as dNBR (difference of the Normalized Burn Ratio) or relative differenced Normalized Burn Ratio (RdNBR), and classified according to the ground reference values of the CBI (Composite Burn Index). UAVs equipped with lidar or hyperspatial sensors allowed the rapid burn severity assessments and further research is needed to utilize the hyperspatial data provided from UAV sensors. Also, lidar technology derived from UAVs are able to provide forest structure parameters and become increasingly available. Similar erosion assessments can improve with the use of UAVs since they are able to provide higher spatial data. The use of remote sensing to analyze vegetation recovery is expected to grow even further in application and prominence as new sensors become available (i.e., UAVs and new satellites) and bring enhanced

spatial, spectral and temporal resolutions to the observations [239]. Current and future trends are the combination of remote sensing data from passive and active sensors either spaceborne or airborne. With the availability of several spaceborne sensors (passive and active) and the emerging improvements of UAVs technologies (flight endurance, sensors, affordability), the development of algorithms that can monitor changes through time irrespective of the characteristics of each platform [279] will improve the overall accuracy of the assessments in almost all the reviewed topics. Finally, recent advances in artificial intelligence (AI) will play a growing role in geospatial technologies as remote sensing (RS) applications [280–282].

Author details

Georgios Zagalikis
Independent Scientist, Ioannina, Greece

*Address all correspondence to: zagalikis@gmail.com

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Food and Agriculture Organization. United Nations Environment Programme. Forests, Biodiversity and People. The State of the World's Forests 2020 [Internet]. 2020 [cited 2023 Jan 17]. pp. 227-49. Available from: www.fao.org/publications%0Ahttp://www.fao.org/documents/card/en/c/ca8642en
- [2] Intergovernmental Panel on Climate Change (IPCC). Technical summary. In: Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press; 2023. pp. 37-118. DOI: 10.1017/9781009325844.002
- [3] Kurvits T, Popescu A, Paulson A, Sullivan A, Ganz D, Burton C, et al. Spreading Like Wildfire: The Rising Threat of Extraordinary Landscape Fires. Nairobi: United Nations Environment Programme; 2022. Available from: <https://policycommons.net/artifacts/2390241/spreading-like-wildfire/3411461/>
- [4] San-Miguel-Ayanz J, Camia A. Forest fires. In 'mapping the impacts of natural hazards and technological accidents in Europe: An overview of the last decade'. European Environment Agency Technical Report N. 2010;13:47-53
- [5] Pyne S. Eternal flame: An introduction to the fire history of the mediterranean. In: Earth Observation of Wildland Fires in Mediterranean Ecosystems. Berlin, Heidelberg: Springer; 2009. pp. 11-26. DOI: 10.1007/978-3-642-01754-4_2
- [6] Amraoui M, Liberato MLR, Calado TJ, DaCamara CC, Coelho LP, Trigo RM, et al. Fire activity over Mediterranean Europe based on information from Meteosat-8. Forest Ecology and Management. 2013;294:62-75
- [7] Pereira MG, Calado TJ, DaCamara CC, Calheiros T. Effects of regional climate change on rural fires in Portugal. Climate Research. 2013;57:187-200 [Internet] [cited 2022 Nov 29]. Available from: <https://www.int-res.com/abstracts/cr/v57/n3/p187-200/>
- [8] Ruffault J, Moron V, Trigo RM, Curt T. Daily synoptic conditions associated with large fire occurrence in Mediterranean France: Evidence for a wind-driven fire regime. International Journal of Climatology. 2017;37:524-533 [Internet] [cited 2023 Jan 17]. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1002/joc.4680>
- [9] Hardy CC. Wildland fire hazard and risk: Problems, definitions, and context. Forest Ecology and Management. 2005;211:73-82
- [10] Chuvieco E, Aguado I, Yebra M, Nieto H, Salas J, Martín MP, et al. Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. Ecological Modelling. 2010;221:46-58
- [11] Fischer MA, di Bella CM, Jobbágy EG. Influence of fuel conditions on the occurrence, propagation and duration of wildland fires: A regional approach. Journal of Arid Environments. 2015;120:63-71
- [12] Calle A, Casanova JL. Forest fires and remote sensing. In: Coskun HG, Cigizoglu HK, Maktav MD, editors. Integration of Information for Environmental Security. NATO

Science for Peace and Security
Series C: Environmental Security.
Dordrecht: Springer; 2008. pp. 261-304.
DOI: 10.1007/978-1-4020-6575-0_19

[13] San-Miguel-Ayanz J, Carlson JD, Alexander M, Tolhurst K, Morgan G, Sneeuwjagt R, et al. Current methods to assess fire danger potential. In: Chuvieco E, editor. *Wildland Fire Danger Estimation and Mapping - The Role of Remote Sensing Data*. Singapore: World Scientific Publishing Co. Pte. Ltd.; 2003. pp. 21-61, 44 p

[14] Gerdzheva AA. A comparative analysis of different wildfire risk assessment models (a case study for Smolyan district, Bulgaria). *European Journal of Geography*. 2014;5:22-36

[15] Adab H, Kanniah KD, Solaimani K. Modeling forest fire risk in the northeast of Iran using remote sensing and GIS techniques. *Natura*. 2013;65:1723-1743 [Internet] [cited 2022 Dec 4]. Available from: <https://link.springer.com/article/10.1007/s11069-012-0450-8>

[16] Gabban A, San-Miguel-Ayanz J, Viegas DX. A comparative analysis of the use of NOAA-AVHRR NDVI and FWI data for forest fire risk assessment. *International Journal of Remote Sensing*. 2008;29:5677-5687

[17] Szpakowski D, Jensen J. A review of the applications of remote sensing in fire ecology. *Remote Sensing (MDPI AG)*. 2019;11:2638 [Internet]. DOI: 10.3390/rs11222638

[18] Finney MA. The challenge of quantitative risk analysis for wildland fire. *Forest Ecology and Management*. 2005;211:97-108

[19] Hernandez C, Drobinski P, Turquety S. How much does weather control fire size and intensity in

the Mediterranean region? *Annales de Geophysique*. 2015;33:931-939 [Internet] [cited 2022 Dec 4]. Available from: <https://angeo.copernicus.org/articles/33/931/2015/>

[20] Nesterov VG. *Combustibility of the Forest and Methods for its Determination*. Moscow, Russia: Goslesbumizdat, USSR State Industry Press; 1949. 76 p

[21] McArthur AG. Fire behaviour in eucalypt forests. *Forestry and Timber Bureau*. 1967;26:26

[22] Deeming JE, Burgan RE, Cohen JD. *The National Fire-Danger Rating System*. Gen. Tech. Rep. INT-GTR-39. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station; 1978. 63 p

[23] van Wagner CE. *Development and Structure of the Canadian Forest Fire Weather Index System*. Forestry Technical Report 35. Ottawa: Canadian Forestry Service, Headquarters; 1987. 35 p

[24] Yu B, Chen F, Li B, Wang L, Wu M. Fire risk prediction using remote sensed products: A case of Cambodia. *Photogramm Eng Remote Sensing*. 2017;83:19-25

[25] Bax V, Francesconi W. Environmental predictors of forest change: An analysis of natural predisposition to deforestation in the tropical Andes region, Peru. *Applied Geography*. 2018;91:99-110

[26] Kocher SD, Butsic V, Smith AMS, Lutz J, Hoffman C, Williamson G, et al. *Governance of Land Use Planning to Reduce Fire Risk to Homes Mediterranean France and California*. 2017. Available from: www.mdpi.com/journal/land

[27] Hong H, Tsangaratos P, Ilia I, Liu J, Zhu AX, Xu C. Applying genetic

algorithms to set the optimal combination of forest fire related variables and model forest fire susceptibility based on data mining models. The case of Dayu County, China. *Science of The Total Environment*. 2018;**630**:1044-1056

[28] de Leeuw J, Vrieling A, Shee A, Atzberger C, Hadgu KM, Biradar CM, et al. The potential and uptake of remote sensing in insurance: A review. *Remote Sensing*. 2014;**6**:10888-10912

[29] Leblon B, Bourgeau-Chavez L, San-Miguel-Ayanz J, Leblon B, Bourgeau-Chavez L, San-Miguel-Ayanz J. Use of remote sensing in wildfire management. Current international perspectives on wildland fires. In: Curkovic S, editor. *Sustainable Development - Authoritative and Leading Edge Content for Environmental Management*. IntechOpen; 2012

[30] Laurin GV, Avezzano R, Bacciu V, del Frate F, Papale D, Virelli M. Cosmo-skymed potential to detect and monitor mediterranean maquis fires and regrowth: A pilot study in capo Figari, Sardinia, Italy. *iForest*. 2018;**11**:389-395

[31] Arroyo LA, Pascual C, Manzanera JA. Fire models and methods to map fuel types: The role of remote sensing. *Forest Ecology and Management*. 2008;**256**:1239-1252

[32] Allgöwer B, Carlson JD, van Wagtenonk JW. Introduction to fire danger rating and remote sensing — Will remote sensing enhance wildland fire danger rating? In: Chuvieco E, editor. *Wildland Fire Danger Estimation and Mapping: The Role of Remote Sensing Data*. Chapter: 1. World Scientific; 2003

[33] Chuvieco E, Riaño D, van Wagtenok J, Morsdof F. Fuel Loads and Fuel Type Mapping. 2003. pp. 119-42

[34] Chuvieco E, Kasischke ES. Remote sensing information for fire management and fire effects assessment. *Journal of Geophysical Research – Biogeosciences*. 2007;**112**:G01S90. DOI: 10.1029/2006JG000230

[35] Vadrevu KP, Eaturu A, Badarinath KVS. Fire risk evaluation using multicriteria analysis—A case study. *Environmental Monitoring and Assessment*. 2009;**166**:223-239 [Internet] [cited 2022 Dec 9]. Available from: <https://link.springer.com/article/10.1007/s10661-009-0997-3>

[36] Qin R, Liu T. A review of landcover classification with very-high resolution remotely sensed optical images—Analysis unit, model scalability and transferability. *Remote Sensing*. 2022;**14**(3):646. DOI: 10.3390/RS14030646

[37] Yankovich KS, Yankovich EP, Baranovskiy NV. Classification of vegetation to estimate forest fire danger using Landsat 8 images: Case study. *Mathematical Problems in Engineering*. 2019;**2019**:1-14. DOI: 10.1155/2019/6296417

[38] Shaik RU, Giovanni L, Fusilli L. New approach of sample generation and classification for wildfire fuel mapping on hyperspectral (prisma) image. In: *IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Brussels, Belgium. 2021. pp. 5417-5420. DOI: 10.1109/IGARSS47720.2021.9554652

[39] Pizaña JMG, Romero JMNH, NC. Remote sensing-based biomass estimation. In: Marghany M, editor. *Environmental Applications of Remote Sensing* [Internet]. Rijeka: IntechOpen; 2016. Ch. 1. DOI: 10.5772/61813

[40] Lu D. The potential and challenge of remote sensing-based biomass

estimation. *International Journal of Remote Sensing*. 2006;**27**:1297-1328

[41] Brandis K, Jacobson C. Estimation of vegetative fuel loads using Landsat TM imagery in New South Wales, Australia. *International Journal of Wildland Fire*. 2003;**12**:185-194 [Internet] [cited 2022 Dec 13]. Available from: <https://www.publish.csiro.au/wf/wf03032>

[42] Reich RM, Lundquist JE, Bravo VA, Reich RM, Lundquist JE, Bravo VA. Spatial models for estimating fuel loads in the Black Hills, South Dakota, USA. *International Journal of Wildland Fire*. 2004;**13**:119-129 [Internet] [cited 2022 Dec 13]. Available from: <https://www.publish.csiro.au/wf/WF02049>

[43] Alonso-Benito A, Arroyo LA, Arbelo M, Hernández-Leal P, González-Calvo A, Alonso-Benito A, et al. Pixel and object-based classification approaches for mapping forest fuel types in Tenerife Island from ASTER data. *International Journal of Wildland Fire*. 2012;**22**:306-317 [Internet] [cited 2022 Dec 13]. Available from: <https://www.publish.csiro.au/wf/WF11068>

[44] Bajocco S, Dragoz E, Gitas I, Smiraglia D, Salvati L, Ricotta C. Mapping Forest fuels through vegetation phenology: The role of coarse-resolution satellite time-series. *PLoS One*. 2015;**10**:e0119811 [Internet] [cited 2022 Dec 13]. Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0119811>

[45] Mallinis G, Galidaki G, Gitas I. A comparative analysis of EO-1 Hyperion, Quickbird and Landsat TM imagery for fuel type mapping of a typical Mediterranean landscape. *Remote Sensing*. 2014;**6**:1684-1704 [Internet] [cited 2022 Dec 13]. Available from: <https://www.mdpi.com/2072-4292/6/2/1684/htm>

[46] Jia GJ, Burke IC, Goetz AFH, Kaufmann MR, Kindel BC. Assessing spatial patterns of forest fuel using AVIRIS data. *Remote Sensing of Environment*. 2006;**102**:318-327

[47] Lasaponara R, Lanorte A, Pignatti S. Multiscale fuel type mapping in fragmented ecosystems: Preliminary results from hyperspectral MIVIS and multispectral Landsat TM data. *International Journal of Remote Sensing*. 2006;**27**:587-593

[48] Arellano-Pérez S, Castedo-Dorado F, López-Sánchez C, González-Ferreiro E, Yang Z, Díaz-Varela R, et al. Potential of Sentinel-2A data to model surface and canopy fuel characteristics in relation to crown fire hazard. *Remote Sensing*. 2018;**10**:1645. DOI: 10.3390/rs10101645

[49] Alonso-Benito A, Arroyo L, Arbelo M, Hernández-Leal P. Fusion of WorldView-2 and LiDAR data to map fuel types in the Canary Islands. *Remote Sensing (MDPI AG)*. 2016;**8**:669 [Internet]. DOI: 10.3390/rs8080669

[50] Arroyo LA, Healey SP, Cohen WB, Cocero D, Manzanera JA. Using object-oriented classification and high-resolution imagery to map fuel types in a Mediterranean region. *Journal of Geophysical Research*. 2006;**111**:4-04 [Internet]. Available from: <http://www.firegrowthmodel>

[51] Keane RE, Burgan R, van Wagten J. Mapping wildland fuels for fire management across multiple scales: Integrating remote sensing, GIS, and biophysical modeling. *International Journal of Wildland Fire*. 2001;**10**:301-319 [Internet] [cited 2022 Dec 9]. Available from: <https://www.publish.csiro.au/wf/wf01028>

[52] García M, Popescu S, Riaño D, Zhao K, Neuenschwander A, Agca M,

- et al. Characterization of canopy fuels using ICESat/GLAS data. *Remote Sensing of Environment*. 2012;**123**:81-89
- [53] Perko R, Raggam H, Deutscher J, Gutjahr K, Schardt M. Forest assessment using high resolution SAR data in X-band. *Remote Sensing*. 2011;**3**:792-815 [Internet] [cited 2013 Dec 30]. Available from: <http://www.mdpi.com/2072-4292/3/4/792/>
- [54] D'este M, Elia M, Giannico V, Spano G, Laforteza R, Sanesi G. Machine learning techniques for fine dead fuel load estimation using multi-source remote sensing data. *Remote Sensing*. 2021;**13**:1658 [Internet] [cited 2022 Dec 10]. Available from: <https://www.mdpi.com/2072-4292/13/9/1658/htm>
- [55] Chuvieco E, Congalton RG. Application of remote sensing and geographic information systems to forest fire hazard mapping. *Remote Sensing of Environment*. 1989;**29**:147-159
- [56] Pimont F, Dupuy J-L, Linn RR, Pimont F, Dupuy J-L, Linn RR. Coupled slope and wind effects on fire spread with influences of fire size: A numerical study using FIRETEC. *International Journal of Wildland Fire*. 2012;**21**:828-842 [Internet] [cited 2022 Dec 16]. Available from: <https://www.publish.csiro.au/wf/WF11122>
- [57] Abdollahi M, Islam T, Gupta A, Hassan Q. An advanced forest fire danger forecasting system: Integration of remote sensing and historical sources of ignition data. *Remote Sensing (MDPI AG)*. 2018;**10**:923 [Internet]. DOI: 10.3390/rs10060923
- [58] Vilar L, Camia A, San-Miguel-Ayanz J. A comparison of remote sensing products and forest fire statistics for improving fire information in mediterranean Europe. *European Journal of Remote Sensing*. 2015;**48**:345-364
- [59] Arino O, Plummer S, Defrenne D, Arino O, Plummer S, Defrenne D. Fire disturbance: The ten years time series of the ATSR world fire atlas. *ESASP*. 2005;**597**:30.1 [Internet] [cited 2022 Dec 17]. Available from: <https://ui.adsabs.harvard.edu/abs/2005ESASP.597E..30A/abstract>
- [60] Giglio L, Descloitres J, Justice CO, Kaufman YJ. An enhanced contextual fire detection algorithm for MODIS. *Remote Sensing of Environment*. 2003;**87**:273-282
- [61] Trigo IF, Dacamara CC, Viterbo P, Roujean JL, Olesen F, Barroso C, et al. The satellite application facility for land surface analysis. *International Journal of Remote Sensing*. 2011;**32**:2725-2744
- [62] Shahbazi M, Théau J, Ménard P. Recent applications of unmanned aerial imagery in natural resource management. *GIScience & Remote Sensing*. 2014;**51**:339-365
- [63] Colomina I, Molina P. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2014;**92**:79-97
- [64] Lechner AM, Foody GM, Boyd DS. Applications in remote sensing to Forest ecology and management. *One Earth* [Internet]. 2020;**2**:405-412. DOI: 10.1016/j.oneear.2020.05.001
- [65] Tang L, Shao G. Drone remote sensing for forestry research and practices. *Journal of Forestry Research*. 2015;**26**:791-797. DOI: 10.1007/s11676-015-0088-y
- [66] Guimarães N, Pádua L, Marques P, Silva N, Peres E, Sousa JJ. Forestry remote sensing from unmanned aerial vehicles: A review focusing on the data, processing and potentialities. *Remote Sensing*

- (MDPI AG). 2020;**12**:1046 [Internet]. DOI: 10.3390/rs12061046
- [67] Manfreda S, McCabe M, Miller P, Lucas R, Pajuelo Madrigal V, Mallinis G, et al. On the use of unmanned aerial systems for environmental monitoring. *Remote Sensing* (MDPI AG). 2018;**10**:641 [Internet]. DOI: 10.3390/rs10040641
- [68] Atole RR, Bello LC, Lirag JR. Eyes in the sky: A review of civilian unmanned aerial vehicles (UAVs). *International Journal of Computers and Applications*. 2017;**173**:36-41
- [69] Dainelli R, Toscano P, di Gennaro SF, Matese A. Recent advances in unmanned aerial vehicle Forest remote sensing—A systematic review. Part I: A general framework. *Forests*. 2021;**12**:327 [Internet] [cited 2022 Dec 20]. Available from: <https://www.mdpi.com/1999-4907/12/3/327/htm>
- [70] Vilar L, Camia A, San-Miguel-Ayanz J. Modelling socio-economic drivers of forest fires in the Mediterranean Europe. *Advances in Forest Fire Research*. 2014:1874-1882. DOI: 10.14195/978-989-26-0884-6_209
- [71] Catry FX, Rego FC, Bação FL, Moreira F. Modeling and mapping wildfire ignition risk in Portugal. *International Journal of Wildland Fire*. 2009;**18**:921-931
- [72] Romero-Calcerrada R, Novillo CJ, Millington JDA, Gomez-Jimenez I. GIS analysis of spatial patterns of human-caused wildfire ignition risk in the SW of Madrid (Central Spain). *Landscape Ecology*. 2008;**23**:341-354
- [73] Arab A, Khodaei A, Eskandarpour R, Thompson MP, Wei Y. Three lines of Defense for wildfire risk Management in Electric Power Grids: A review. *IEEE Access*. 2021;**9**:61577-61593
- [74] Xu K, Zhang X, Chen Z, Wu W, Li T. Risk assessment for wildfire occurrence in high-voltage power line corridors by using remote-sensing techniques: A case study in Hubei Province, China. *International Journal of Remote Sensing*. 2016;**37**:4818-4837
- [75] Park A, Rajabi F, Weber R. Slash or Burn: Power Line and Vegetation Classification for Wildfire Prevention [Internet]. 2021. Available from: <http://arxiv.org/abs/2105.03804>
- [76] Su C, Wu X, Guo Y, Lai CS, Xu L, Zhao X. Automatic multi-source data fusion technique of powerline corridor using UAV Lidar. In: 2022 IEEE International Smart Cities Conference (ISC2), Pafos, Cyprus. 2022. pp. 1-5. DOI: 10.1109/ISC255366.2022.9922293
- [77] Ghorbanzadeh O, Blaschke T, Gholamnia K, Aryal J. Forest fire susceptibility and risk mapping using social/infrastructural vulnerability and environmental variables. *Fire*. 2019;**2**:1-27
- [78] Pourghasemi H, Beheshtirad M, Pradhan B. A comparative assessment of prediction capabilities of modified analytical hierarchy process (M-AHP) and Mamdani fuzzy logic models using Netcad-GIS for forest fire susceptibility mapping. *Geomatics, Natural Hazards and Risk*. 2016;**7**:861-885
- [79] Lamat R, Kumar M, Kundu A, Lal D. Forest fire risk mapping using analytical hierarchy process (AHP) and earth observation datasets: A case study in the mountainous terrain of Northeast India. *SN Applied Sciences*. 2021;**3**:1-15 [Internet] [cited 2022 Dec 26]. Available from: <https://link.springer.com/article/10.1007/s42452-021-04391-0>
- [80] Serda M, Becker FG, Cleary M, Team RM, Holtermann H, The D, et al. Wildfire susceptibility evaluation by

- integrating an analytical network process approach into GIS-based analyses. G. Balint, Antala B, Carty C, Mabieme J-MA, Amar IB, Kaplanova A, editors. *International Journal on Advanced Science, Engineering and Technology*. 2018;7:48-53 [Internet] [cited 2022 Dec 26]. Available from: <https://uni-salzburg.eu/elsevierpure.com/en/publications/wildfire-susceptibility-evaluation-by-integrating-an-analytical-n-2>
- [81] Mohajane M, Costache R, Karimi F, Bao Pham Q, Essahlaoui A, Nguyen H, et al. Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area. *Ecological Indicators*. 2021;129:107869 [Internet] [cited 2022 Dec 16]. DOI: 10.1016/j.ecolind.2021.107869
- [82] Karpatne A, Ebert-Uphoff I, Ravela S, Babaie HA, Kumar V. Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering*. 2017;31:1544-1554 [Internet] [cited 2022 Dec 28]. Available from: <https://arxiv.org/abs/1711.04708v1>
- [83] Rishickesh R, Shahina A, Nayeemulla KA. Predicting forest fires using supervised and ensemble machine learning algorithms. *International Journal of Recent Technology and Engineering*. 2019;8:3697-3705
- [84] Dutta R, Das A, Aryal J. Big data integration shows Australian bush-fire frequency is increasing significantly. *Royal Society Open Science*. 2016;3(2). DOI: 10.1098/RSOS.150241
- [85] de Souza FT, Koerner TC, Chlad R. A data-based model for predicting wildfires in Chapada das mesas National Park in the state of Maranhão. *Environ. Earth Science*. 2015;74:3603-3611 [Internet] [cited 2022 Dec 26]. Available from: <https://link.springer.com/article/10.1007/s12665-015-4421-8>
- [86] Kim SJ, Lim C-H, Kim G, Lee J, Geiger T, Rahmati O, et al. Multi-temporal analysis of Forest fire probability using socio-economic and environmental variables. *Remote Sensing*. 2019;11:86
- [87] Oliveira S, Oehler F, San-Miguel-Ayanz J, Camia A, Pereira JMC. Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random Forest. *Forest Ecology and Management*. 2012;275:117-129 [Internet] [cited 2022 Dec 26]. Available from: https://www.researchgate.net/publication/236781184_Modeling_spatial_patterns_of_fire_occurrence_in_Mediterranean_Europe_using_Multiple_Regression_and_Random_Forest
- [88] Miller RK, Field CB, Mach KJ. Factors influencing adoption and rejection of fire hazard severity zone maps in California. *International Journal of Disaster Risk Reduction*. 2020;50. DOI: 10.1016/J.IJDRR.2020.101686
- [89] Chuvieco Emilio. Wildland fire danger: estimation and mapping : the role of remote sensing data. 2003 [cited 2022 Dec 28]. pp. 264. Available from: https://books.google.com/books/about/Wildland_Fire_Danger_Estimation_and_Mapp.html?hl=el&id=a2LF_1a8IPcC
- [90] Rothermel RC, Wilson RA, Morris GA, Sackett SS. Modeling Moisture Content of Fine Dead Wildland Fuels: Input to the BEHAVE Fire Prediction System. Research Paper INT-359. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station; 1986. 61 p. DOI: 10.2737/INT-RP-359
- [91] Piñol J, Filella I, Ogaya R, Peñuelas J. Ground-based spectroradiometric estimation of live fine fuel moisture of Mediterranean plants. *Agricultural and Forest Meteorology*. 1998;90:173-186

- [92] Chuvieco E, Deshayes M, Stach N, Cocero D, Riaño D. Short-term fire risk: Foliage moisture content estimation from satellite data. *Remote Sensing of Large Wildfires*. 1999;17-38. DOI: 10.1007/978-3-642-60164-4_3
- [93] Luo K, Quan X, He B, Yebra M. Effects of live fuel moisture content on wildfire occurrence in fire-prone regions over Southwest China. *Forests (MDPI AG)*. 2019;10:887 [Internet]. DOI: 10.3390/f10100887
- [94] Nolan RH, Boer MM, Resco De Dios V, Caccamo G, Bradstock RA. Large-scale, dynamic transformations in fuel moisture drive wildfire activity across southeastern Australia. *Geophysical Research Letters*. 2016;43:4229-4238 [Internet] [cited 2022 Dec 28]. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1002/2016GL068614>
- [95] Chuvieco E, Aguado I, Salas J, García M, Yebra M, Oliva P. Satellite remote sensing contributions to wildland fire science and management. *Current Forestry Reports*. 2020;6:81-96 [Internet] [cited 2022 Dec 28]. Available from: <https://link.springer.com/article/10.1007/s40725-020-00116-5>
- [96] Boer MM, Nolan RH, Resco De Dios V, Clarke H, Price OF, Bradstock RA. Changing weather extremes call for early warning of potential for catastrophic fire. *Earths. Futures*. 2017;5:1196-1202 [Internet] [cited 2022 Dec 28]. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1002/2017EF000657>
- [97] Caccamo G, Chisholm LA, Bradstock RA, Puotinen ML, Phippen BG, Caccamo G, et al. Monitoring live fuel moisture content of heathland, shrubland and sclerophyll forest in South-Eastern Australia using MODIS data. *International Journal of Wildland Fire*. 2011;21:257-269 [Internet] [cited 2022 Dec 29]. Available from: <https://www.publish.csiro.au/wf/WF11024>
- [98] Argañaraz JP, Landi MA, Bravo SJ, Gavier-Pizarro GI, Scavuzzo CM, Bellis LM. Estimation of live fuel moisture content from MODIS images for fire danger assessment in southern Gran Chaco. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2016;9:5339-5349
- [99] Jurdao S, Yebra M, Guerschman JP, Chuvieco E. Regional estimation of woodland moisture content by inverting radiative transfer models. *Remote Sensing of Environment*. 2013;132:59-70
- [100] Yebra M, Quan X, Riaño D, Rozas Larraondo P, van Dijk AIJM, Cary GJ. A fuel moisture content and flammability monitoring methodology for continental Australia based on optical remote sensing. *Remote Sensing of Environment*. 2018;212:260-272
- [101] Zhu L, Webb GI, Yebra M, Scortechini G, Miller L, Petitjean F. Live fuel moisture content estimation from MODIS: A deep learning approach. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2021;179:81-91
- [102] Rao K, Williams AP, Flefil JF, Konings AG. SAR-enhanced mapping of live fuel moisture content. *Remote Sensing of Environment*. 2020;245:111797
- [103] Marino E, Yebra M, Guillén-Climent M, Algeet N, Tomé JL, Madrigal J, et al. Investigating live fuel moisture content estimation in fire-prone shrubland from remote sensing using empirical modelling and RTM simulations. *Remote Sensing (MDPI AG)*. 2020;12:2251 [Internet]. DOI: 10.3390/rs12142251

- [104] Hunt ER, Rock BN. Detection of changes in leaf water content using near- and middle-infrared reflectances. *Remote Sensing of Environment*. 1989;**30**:43-54
- [105] Seelig HD, Hoehn A, Stodieck LS, Klaus DM, Adams WW, Emery WJ. The assessment of leaf water content using leaf reflectance ratios in the visible, near-, and short-wave-infrared. *International Journal of Remote Sensing*. 2008;**29**:3701-3713
- [106] Cunill Camprubí À, González-Moreno P, de Dios VR. Live fuel moisture content mapping in the Mediterranean Basin using random forests and combining MODIS spectral and thermal data. *Remote Sensing (MDPI AG)*. 2022;**14**:3162 [Internet]. DOI: 10.3390/rs14133162
- [107] Costa-Saura JM, Balaguer-Beser Á, Ruiz LA, Pardo-Pascual JE, Soriano-Sancho JL. Empirical models for spatio-temporal live fuel moisture content estimation in mixed Mediterranean vegetation areas using sentinel-2 indices and meteorological data. *Remote Sensing (MDPI AG)*. 2021;**13**:3726 [Internet]. DOI: 10.3390/rs13183726
- [108] Shu Q, Quan X, Yebra M, Liu X, Wang L, Zhang Y. Evaluating the Sentinel-2A satellite data for fuel moisture content retrieval. In: *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan*. 2019. pp. 9416-9419. DOI: 10.1109/IGARSS.2019.8900104
- [109] Fan L, Wigneron JP, Xiao Q, Al-Yaari A, Wen J, Martin-StPaul N, et al. Evaluation of microwave remote sensing for monitoring live fuel moisture content in the Mediterranean region. *Remote Sensing of Environment*. 2018;**205**:210-223
- [110] Wang L, Quan X, He B, Yebra M, Xing M, Liu X. Assessment of the dual Polarimetric sentinel-1A data for Forest fuel moisture content estimation. *Remote Sensing*. 2019;**11**:1568 [Internet] [cited 2022 Dec 30]. Available from: <https://www.mdpi.com/2072-4292/11/13/1568/htm>
- [111] Yebra M, Dennison PE, Chuvieco E, Riaño D, Zylstra P, Hunt ER, et al. A global review of remote sensing of live fuel moisture content for fire danger assessment: Moving towards operational products. *Remote Sensing of Environment*. 2013;**136**:455-468
- [112] Alexander ME, Cole FV. Rating fire danger in Alaska ecosystems: CFFDRS provides an invaluable guide to systematically evaluating burning conditions. *Fireline (Fort Wainwright, Alaska: USDI Bureau of Land Management)*. 2001;**12**(4):2-3
- [113] Groot WJD, Field RD, Brady MA, Roswintiarti O, Mohamad M. Development of the Indonesian and Malaysian fire danger rating systems. *Mitigation and Adaptation Strategies for Global Change*. 2007;**12**:165-180
- [114] San-Miguel-Ayanz J, Barbosa P, Libertá G, Schmuck G, Schulte E, Bucella P. The European forest fire information system: A European strategy towards forest fire management. In: *Proceedings of the 3rd International Wildland Fire Conference, Sydney, Australia*. Washington, DC: US Department of the Interior, Bureau of Land Management CD-ROM; 2003
- [115] Viegas DX, Bovio G, Ferreira A, Nosenzo A, Sol B. Comparative study of various methods of fire danger evaluation in southern Europe. *International Journal of Wildland Fire*. 2000;**9**:235-246
- [116] San-Miguel-Ayanz J, Schulte E, Schmuck G, Camia A, Strobl P,

- Liberta G, et al. Comprehensive monitoring of wildfires in Europe: The European forest fire information system (EFFIS). In: Tiefenbacher J, editor. *Approaches to Managing Disaster - Assessing Hazards, Emergencies and Disaster Impacts*. 2012
- [117] Chowdhury EH, Hassan QK. Development of a new daily-scale forest fire danger forecasting system using remote sensing data. *Remote Sensing*. 2015;7:2431-2448
- [118] Sirca C, Salis M, Arca B, Duce P, Spano D. Assessing the performance of fire danger indexes in a Mediterranean area. *iForest*. 2018;11:563 [Internet] [cited 2022 Dec 9]. Available from: <https://iforest.sisef.org/abstract/?id=ifor2679-011>
- [119] Fiorucci P, Gaetani F, Minciardi R. Development and application of a system for dynamic wildfire risk assessment in Italy. *Environmental Modelling & Software*. 2008;23:690-702
- [120] Laneve G, Pampanoni V, Shaik RU. The daily fire Hazard index: A fire danger rating method for Mediterranean areas. *Remote Sensing*. 2020;12:2356 [Internet] [cited 2022 Dec 31]. Available from: <https://www.mdpi.com/2072-4292/12/15/2356/htm>
- [121] Liu X, He B, Quan X, Yebra M, Qiu S, Yin C, et al. Near real-time extracting wildfire spread rate from Himawari-8 satellite data. *Remote Sensing*. 2018;10:1-15
- [122] Ying L, Shen Z, Yang M, Piao S. Wildfire detection probability of MODIS fire products under the constraint of environmental factors: A study based on confirmed ground wildfire records. *Remote Sensing (MDPI AG)*. 2019;11:3031 [Internet]. DOI: 10.3390/rs11243031
- [123] Lin Z, Chen F, Niu Z, Li B, Yu B, Jia H, et al. An active fire detection algorithm based on multi-temporal FengYun-3C VIRR data. *Remote Sensing of Environment*. 2018;211:376-387
- [124] Flannigan MD, Vonder Haar TH. Forest fire monitoring using NOAA satellite AVHRR. *Canadian Journal of Forest Research*. 1986;16:975-982
- [125] Justice CO, Korontzi SA. A review of satellite fire monitoring and the requirements for global environmental change research. In: Ahem F, Goldammer G, Justice CO, editors. *Global and Regional Vegetation Fire Monitoring From Space: Planning a Coordinated International Effort*. The Hague, The Netherlands: SPB Academic Publishing; 2001. pp. 1-18
- [126] Giglio L, Csiszar I, Justice CO. Global distribution and seasonality of active fires as observed with the Terra and Aqua moderate resolution imaging spectroradiometer (MODIS) sensors. *Journal of Geophysical Research: Biogeosciences*. 2006;111(2). DOI: 10.1029/2005JG000142
- [127] Schroeder W, Prins E, Giglio L, Csiszar I, Schmidt C, Morisette J, et al. Validation of GOES and MODIS active fire detection products using ASTER and ETM+ data. *Remote Sensing of Environment*. 2008;112:2711-2726
- [128] Giglio L, Schroeder W, Justice CO. The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sensing of Environment*. 2016;178:31-41
- [129] Morisette JT, Giglio L, Csiszar I, Justice CO. Validation of the MODIS active fire product over southern Africa with ASTER data. *International Journal of Remote Sensing*. 2005;26:4239-4264

- [130] Forghani A, Reddy S, Thankappan M, Cechet B. Validation of MODIS and AVHRR fire detections in Australia. *International Journal of Geoinformatics*. 2021;17:117-131
- [131] Cahoon DR, Stocks BJ, Alexander ME, Baum BA, Goldammer JG. Wildland Fire Detection from Space: Theory and Application. 2000 [cited 2023 Jan 5]. pp. 151-69. Available from: https://link.springer.com/chapter/10.1007/0-306-47959-1_9
- [132] Maier SW, Russell-Smith J, Edwards AC, Yates C. Sensitivity of the MODIS fire detection algorithm (MOD14) in the savanna region of the Northern Territory, Australia. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2013;76:11-16
- [133] Hawbaker TJ, Radeloff VC, Syphard AD, Zhu Z, Stewart SI. Detection rates of the MODIS active fire product in the United States. *Remote Sensing of Environment*. 2008;112:2656-2664
- [134] Justice CO, Giglio L, Korontzi S, Owens J, Morisette JT, Roy D, et al. The MODIS fire products. *Remote Sensing of Environment*. 2002;83:244-262
- [135] Schroeder W, Oliva P, Giglio L, Quayle B, Lorenz E, Morelli F. Active fire detection using Landsat-8/OLI data. *Remote Sensing of Environment*. 2016;185:210-220
- [136] Gargiulo M, Dell' Aglio DAG, Iodice A, Riccio D, Ruello G. A CNN-based super-resolution technique for active fire detection on Sentinel-2 data. In: 2019 Photonics & Electromagnetics Research Symposium - Spring (PIERS-Spring), Rome, Italy. 2019. pp. 418-426. DOI: 10.1109/PIERS-Spring46901.2019.9017857
- [137] Xu G, Zhong X. Real-time wildfire detection and tracking in Australia using geostationary satellite: Himawari-8. *Remote Sensing Letters*. 2017;8:1052-1061
- [138] Maeda N, Tonooka H. Early stage Forest fire detection from Himawari-8 AHI images using a modified MOD14 algorithm combined with machine learning. *Sensors*. 2022;23:210 [Internet] [cited 2023 Jan 5]. Available from: <https://www.mdpi.com/1424-8220/23/1/210>
- [139] Hall JV, Zhang R, Schroeder W, Huang C, Giglio L. Validation of GOES-16 ABI and MSG SEVIRI active fire products. *International Journal of Applied Earth Observation and Geoinformation*. 2019;83:101928
- [140] Wang W, Qu JJ, Hao X, Liu Y, Sommers WT. An improved algorithm for small and cool fire detection using MODIS data: A preliminary study in the southeastern United States. *Remote Sensing of Environment*. 2007;108:163-170
- [141] Zajkowski T. Selection of appropriate class UAS/sensors to support fire monitoring: Experiences in the United States. *Handbook of Unmanned Aerial Vehicles*. 2015:2723-2754. DOI: 10.1007/978-90-481-9707-1_73/COVER
- [142] Nex F, Remondino F. Preface: Latest developments, methodologies, and applications based on UAV platforms. *Drones*. 2019;3:26 [Internet] [cited 2023 Jan 6]. Available from: <https://www.mdpi.com/2504-446X/3/1/26/htm>
- [143] Watts AC, Ambrosia VG, Hinkley EA. Unmanned aircraft systems in remote sensing and scientific research: Classification and considerations of use. *Remote Sensing*. 2012;4:1671-1692

- [144] Yuan C, Liu Z, Zhang Y. UAV-based forest fire detection and tracking using image processing techniques. 2015 international conference on unmanned aircraft systems. ICUAS. 2015;2015:639-643
- [145] Akhloufi MA, Couturier A, Castro NA. Unmanned aerial vehicles for wildland fires: Sensing, perception, cooperation and assistance. Drones. 2021;5:15 [Internet] [cited 2023 Jan 6]. Available from: <https://www.mdpi.com/2504-446X/5/1/15/htm>
- [146] Randerson JT, Chen Y, van der Werf GR, Rogers BM, Morton DC. Global burned area and biomass burning emissions from small fires. Journal of Geophysical Research – Biogeosciences. 2012;117:4012 [Internet] [cited 2023 Jan 7]. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1029/2012JG002128>
- [147] Giglio L, Loboda T, Roy DP, Quayle B, Justice CO. An active-fire based burned area mapping algorithm for the MODIS sensor. Remote Sensing of Environment. 2009;113:408-420
- [148] Langmann B, Duncan B, Textor C, Trentmann J, van der Werf GR. Vegetation fire emissions and their impact on air pollution and climate. Atmospheric Environment. 2009;43:107-116
- [149] Yang W, Jiang X. High-resolution estimation of air pollutant emissions from vegetation burning in China (2000-2018). Frontiers in Environmental Science. 2022;10:1333
- [150] Pereira JMC, Sousa AMO, Sá ACL, Martín MP, Chuvieco E. Regional-scale burnt area mapping in Southern Europe using NOAA-AVHRR 1 km data. Remote Sensing of Large Wildfires. 1999:139-155. DOI: 10.1007/978-3-642-60164-4_8
- [151] Al-Rawi KR, Casanova JL, Calle A. Burned area mapping system and fire detection system, based on neural networks and NOAA-AVHRR imagery. International Journal of Remote Sensing. 2001;22:2015-2032
- [152] Grégoire JM, Tansey K, Silva JMN. The GBA2000 initiative: Developing a global burnt area database from SPOT-VEGETATION imagery. International Journal of Remote Sensing. 2003;24:1369-1376
- [153] Tansey K, Grégoire JM, Stroppiana D, Sousa A, Silva J, Pereira JMC, et al. Vegetation burning in the year 2000: Global burned area estimates from SPOT VEGETATION data. Journal of Geophysical Research: Atmospheres. 2004;109(D14). DOI: 10.1029/2003JD003598
- [154] Simon M, Plummer S, Fierens F, Hoelzemann JJ, Arino O. Burnt area detection at global scale using ATSR-2: The GLOBSCAR products and their qualification. Journal of Geophysical Research: Atmospheres. 2004;109(D14). DOI: 10.1029/2003JD003622
- [155] Roy DP, Jin Y, Lewis PE, Justice CO. Prototyping a global algorithm for systematic fire-affected area mapping using MODIS time series data. Remote Sensing of Environment. 2005;97:137-162
- [156] Boschetti L, Roy D, Barbosa P, Boca R, Justice C. A MODIS assessment of the summer 2007 extent burned in Greece. International Journal of Remote Sensing. 2008;29:2433-2436
- [157] Roy DP, Boschetti L. Southern Africa validation of the MODIS, L3JRC, and GlobCarbon burned-area products. IEEE Transactions on Geoscience and Remote Sensing. 2009;47:1032-1044
- [158] Padilla M, Stehman SV, Ramo R, Corti D, Hantson S, Oliva P, et al. Comparing the accuracies of remote

sensing global burned area products using stratified random sampling and estimation. *Remote Sensing of Environment*. 2015;**160**:114-121

[159] San-Miguel-Ayanz J, Pereira JMC, Boca R, Strobl P, Kucera J, Pekkarinen A. Forest fires in the European Mediterranean region: Mapping and analysis of burned areas. *Earth Observation of Wildland Fires in Mediterranean Ecosystems*. 2009:189-203. DOI: 10.1007/978-3-642-01754-4_13

[160] Roy DP, Wulder MA, Loveland TR, Woodcock CE, Allen RG, Anderson MC, et al. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*. 2014;**145**:154-172

[161] Wang Q, Blackburn GA, Onojeghuo AO, Dash J, Zhou L, Zhang Y, et al. Fusion of Landsat 8 OLI and Sentinel-2 MSI data. *IEEE Transactions on Geoscience and Remote Sensing*. 2017;**55**:3885-3899

[162] Roy DP, Huang H, Boschetti L, Giglio L, Yan L, Zhang HH, et al. Landsat-8 and Sentinel-2 burned area mapping - A combined sensor multi-temporal change detection approach. *Remote Sensing of Environment*. 2019;**231**. DOI: 10.1016/j.rse.2019.111254

[163] Drusch M, del Bello U, Carlier S, Colin O, Fernandez V, Gascon F, et al. Sentinel-2: ESA's optical high-resolution Mission for GMES operational services. *Remote Sensing of Environment*. 2012;**120**:25-36

[164] Filipponi F. Exploitation of Sentinel-2 time series to map burned areas at the national level: A case study on the 2017 Italy wildfires. *Remote Sensing (MDPI AG)*. 2019;**11**:622 [Internet]. DOI: 10.3390/rs11060622

[165] Quintano C, Fernández-Manso A, Fernández-Manso O. Combination of Landsat and Sentinel-2 MSI data for initial assessing of burn severity. *International Journal of Applied Earth Observation and Geoinformation*. 2018;**64**:221-225

[166] Syifa M, Panahi M, Lee C-W. Mapping of post-wildfire burned area using a hybrid algorithm and satellite data: The case of the camp fire wildfire in California, USA. *Remote Sensing (MDPI AG)*. 2020;**12**:623 [Internet]. DOI: 10.3390/rs12040623

[167] Katagis T, Gitas IZ, Mitri GH. An object-based approach for fire history reconstruction by using three generations of landsat sensors. *Remote Sensing*. 2014;**6**:5480-5496

[168] Boschetti L, Roy DP, Justice CO, Humber ML. MODIS–Landsat fusion for large area 30 m burned area mapping. *Remote Sensing of Environment*. 2015;**161**:27-42

[169] Chuvieco E, Mouillot F, van der Werf GR, San Miguel J, Tanasse M, Koutsias N, et al. Historical background and current developments for mapping burned area from satellite earth observation. *Remote Sensing of Environment*. 2019;**225**:45-64

[170] French NHF, Bourgeau-Chavez LL, Wang Y, Kasischke ES. Initial observations of radarsat imagery at fire-disturbed sites in interior Alaska. *Remote Sensing of Environment*. 1999;**68**:89-94

[171] Siegert F, Ruecker G. Use of multitemporal ERS-2 SAR images for identification of burned scars in south-east Asian tropical rainforest. *International Journal of Remote Sensing*. 2000;**21**:831-837

[172] Gimeno M, San-Miguel-Ayanz J. Evaluation of RADARSAT-1 data for

identification of burnt areas in southern Europe. *Remote Sensing of Environment*. 2004;**92**:370-375

[173] Gimeno M, San-Miguel-Ayanz J, Schmuck G. Identification of burnt areas in Mediterranean forest environments from ERS-2 SAR time series. *International Journal of Remote Sensing*. 2004;**25**:4873-4888

[174] de Luca G, Silva JMN, Modica G. Regional-scale burned area mapping in Mediterranean regions based on the multitemporal composite integration of Sentinel-1 and Sentinel-2 data. *GIScience & Remote Sensing*. 2022;**59**:1678-1705 [Internet] [cited 2023 Jan 8]. Available from: <https://www.tandfonline.com/doi/abs/10.1080/15481603.2022.2128251>

[175] Keeley JE. Fire intensity, fire severity and burn severity: A brief review and suggested usage. *International Journal of Wildland Fire*. 2009;**18**:116-126

[176] Benavides-Solorio J, MacDonald LH. Post-fire runoff and erosion from simulated rainfall on small plots, Colorado front range. *Hydrological Processes*. 2001;**15**:2931-2952 [Internet] [cited 2023 Jan 9]. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1002/hyp.383>

[177] Martin DA, Moody JA. Comparison of soil infiltration rates in burned and unburned mountainous watersheds. *Hydrological Processes*. 2001;**15**:2893-2903

[178] Key CH, Benson NC. Landscape assessment (LA) sampling and analysis methods. In: Lutes DC, Keane RE, Caratti JF, Key CH, Benson NC, Sutherland S, et al., editors. FIREMON: Fire Effects Monitoring and Inventory System. Gen. Tech. Rep. RMRS-GTR-164-CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station; 2006. p. LA-1-55

[179] Kasischke ES, Turetsky MR, Ottmar RD, French NHF, Hoy EE, Kane ES, et al. Evaluation of the composite burn index for assessing fire severity in Alaskan black spruce forests. *International Journal of Wildland Fire*. 2008;**17**:515-526 [Internet] [cited 2023 Jan 9]. Available from: <https://www.publish.csiro.au/wf/WF08002>

[180] Chuvieco E, Riaño D, Danson FM, Martin P. Use of a radiative transfer model to simulate the postfire spectral response to burn severity. *Journal of Geophysical Research: Biogeosciences*. 2006;**111**(G4):4-9. DOI: 10.1029/2005JG000143

[181] Kokaly RF, Rockwell BW, Haire SL, King TVV. Characterization of post-fire surface cover, soils, and burn severity at the Cerro Grande fire, New Mexico, using hyperspectral and multispectral remote sensing. *Remote Sensing of Environment*. 2007;**106**:305-325

[182] Roy DP, Boschetti L, Trigg SN. Remote sensing of fire severity: Assessing the performance of the normalized burn ratio. *IEEE Geoscience and Remote Sensing Letters*. 2006;**3**:112-116

[183] Barrett K, Kasischke ES, McGuire AD, Turetsky MR, Kane ES. Modeling fire severity in black spruce stands in the Alaskan boreal forest using spectral and non-spectral geospatial data. *Remote Sensing of Environment*. 2010;**114**:1494-1503

[184] Hoy EE, French NHF, Turetsky MR, Trigg SN, Kasischke ES. Evaluating the potential of Landsat TM/ETM+ imagery for assessing fire severity in Alaskan black spruce forests. *International Journal of Wildland Fire*. 2008;**17**:500-514

[185] Verbyla D, Lord R. Estimating post-fire organic soil depth in the Alaskan

boreal forest using the normalized burn ratio. *International Journal of Remote Sensing*. 2008;**29**:3845-3853

[186] Hall RJ, Freeburn JT, de Groot WJ, Pritchard JM, Lynham TJ, Landry R. Remote sensing of burn severity: Experience from western Canada boreal fires. *International Journal of Wildland Fire*. 2008;**17**:476-489

[187] Veraverbeke S, Verstraeten WW, Lhermitte S, Goossens R, Veraverbeke S, Verstraeten WW, et al. Evaluating Landsat thematic mapper spectral indices for estimating burn severity of the 2007 Peloponnese wildfires in Greece. *International Journal of Wildland Fire*. 2010;**19**:558-569 [Internet] [cited 2023 Jan 9]. Available from: <https://www.publish.csiro.au/wf/WF09069>

[188] Veraverbeke S, Lhermitte S, Verstraeten WW, Goossens R. Evaluation of pre/post-fire differenced spectral indices for assessing burn severity in a mediterranean environment with landsat thematic mapper. *International Journal of Remote Sensing*. 2011;**32**:3521-3537

[189] Miller JD, Thode AE. Quantifying burn severity in a heterogeneous landscape with a relative version of the delta normalized burn ratio (dNBR). *Remote Sensing of Environment*. 2007;**109**:66-80

[190] Soverel NO, Coops NC, Perrakis DDB, Daniels LD, Gergel SE. The transferability of a dNBR-derived model to predict burn severity across 10 wildland fires in western Canada. *International Journal of Wildland Fire*. 2011;**20**:518-531

[191] Wu Z, He HS, Liang Y, Cai L, Lewis BJ. Determining relative contributions of vegetation and topography to burn severity from LANDSAT imagery. *Environmental Management*. 2013;**52**:821-836

[192] Epting J, Verbyla D, Sorbel B. Evaluation of remotely sensed indices for assessing burn severity in interior Alaska using Landsat TM and ETM+. *Remote Sensing of Environment*. 2005;**96**:328-339

[193] Soverel NO, Perrakis DDB, Coops NC. Estimating burn severity from Landsat dNBR and RdNBR indices across western Canada. *Remote Sensing of Environment*. 2010;**114**:1896-1909

[194] Slaton MR, Hunt ER, Smith WK. Estimating near-infrared leaf reflectance from leaf structural characteristics. *American Journal of Botany*. 2001;**88**:278-284 [Internet] [cited 2023 Jan 10]. Available from: <https://onlinelibrary.wiley.com/doi/full/10.2307/2657019>

[195] Quintano C, Fernández-Manso A, Calvo L, Marcos E, Valbuena L. Land surface temperature as potential indicator of burn severity in forest Mediterranean ecosystems. *International Journal of Applied Earth Observation and Geoinformation*. 2015;**36**:1-12

[196] Zheng Z, Zeng Y, Li S, Huang W. A new burn severity index based on land surface temperature and enhanced vegetation index. *International Journal of Applied Earth Observation and Geoinformation*. 2016;**45**:84-94

[197] Veraverbeke S, Verstraeten WW, Lhermitte S, van de Kerchove R, Goossens R, Veraverbeke S, et al. Assessment of post-fire changes in land surface temperature and surface albedo, and their relation with fire–burn severity using multitemporal MODIS imagery. *International Journal of Wildland Fire*. 2012;**21**:243-256 [Internet] [cited 2023 Jan 10]. Available from: <https://www.publish.csiro.au/wf/WF10075>

[198] Horler DNH, Dockray M, Barber J. The red edge of plant leaf reflectance.

International Journal of Remote Sensing. 2007;**4**:273-288 [Internet] [cited 2023 Jan 10]. Available from: <https://www.tandfonline.com/doi/abs/10.1080/01431168308948546>

[199] Fernández-Manso A, Fernández-Manso O, Quintano C. SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity. International Journal of Applied Earth Observation and Geoinformation. 2016;**50**:170-175

[200] Filipponi F. BAIS2: Burned area index for Sentinel-2. Proceedings. 2018;**2**:364 [Internet] [cited 2023 Jan 10]. Available from: <https://www.mdpi.com/2504-3900/2/7/364/htm>

[201] Morgan P, Keane RE, Dillon GK, Jain TB, Hudak AT, Karau EC, et al. Challenges of assessing fire and burn severity using field measures, remote sensing and modelling. International Journal of Wildland Fire. 2014;**23**:1045-1060

[202] Wang C, Glenn NF. Estimation of fire severity using pre- and post-fire LiDAR data in sagebrush steppe rangelands. International Journal of Wildland Fire. 2009;**18**:848-856 [Internet] [cited 2023 Jan 10]. Available from: <https://www.publish.csiro.au/wf/WF08173>

[203] Wulder MA, White JC, Alvarez F, Han T, Rogan J, Hawkes B. Characterizing boreal forest wildfire with multi-temporal Landsat and LIDAR data. Remote Sensing of Environment. 2009;**113**:1540-1555

[204] Fernandez-Manso A, Quintano C, Roberts DA. Burn severity analysis in Mediterranean forests using maximum entropy model trained with EO-1 Hyperion and LiDAR data. ISPRS Journal of Photogrammetry and Remote Sensing. 2019;**155**:102-118

[205] Kane VR, North MP, Lutz JA, Churchill DJ, Roberts SL, Smith DF, et al. Assessing fire effects on forest spatial structure using a fusion of Landsat and airborne LiDAR data in Yosemite National Park. Remote Sensing of Environment. 2014;**151**:89-101

[206] Viedma O, Almeida DRA, Moreno JM. Postfire tree structure from high-resolution LiDAR and RBR sentinel 2A fire severity metrics in a pinus halepensis-dominated burned stand. Remote Sensing. 2020;**12**:1-21

[207] Carvajal-Ramírez F, da Silva JRM, Agüera-Vega F, Martínez-Carricondo P, Serrano J, Moral FJ. Evaluation of fire severity indices based on pre- and post-fire multispectral imagery sensed from UAV. Remote Sensing (MDPI AG). 2019;**11**:993 [Internet]. DOI: 10.3390/rs11090993

[208] McKenna P, Erskine PD, Lechner AM, Phinn S. Measuring fire severity using UAV imagery in semi-arid Central Queensland, Australia. International Journal of Remote Sensing. 2017;**38**:4244-4264 [Internet] [cited 2023 Jan 11]. Available from: <https://www.tandfonline.com/doi/abs/10.1080/01431161.2017.1317942>

[209] Woo H, Acuna M, Madurapperuma B, Jung G, Woo C, Park J. Application of maximum likelihood and spectral angle mapping classification techniques to evaluate forest fire severity from UAV multi-spectral images in South Korea. Sensors and Materials. 2021;**33**:3745-3760

[210] Qi Y, Coops NC, Daniels LD, Butson CR. Assessing the effects of burn severity on post-fire tree structures using the fused drone and mobile laser scanning point clouds. Frontiers in Environmental Science. 2022;**10**:949442. DOI: 10.3389/FENVS.2022.949442/BIBTEX

- [211] Ye J, Cui Z, Zhao F, Liu Q. Automated extraction of Forest burn severity based on light and small UAV visible remote sensing images. *Forests*. 2022;**13**
- [212] Fraser RH, van der Sluijs J, Hall RJ. Calibrating satellite-based indices of burn severity from UAV-derived metrics of a burned boreal forest in NWT, Canada. *Remote Sensing*. 2017;**9**
- [213] Shin J, Seo W, Kim T, Park J, Woo C. Using UAV multispectral images for classification of forest burn severity-A case study of the 2019 Gangneung forest fire. *Forests (MDPI AG)*. 2019;**10**:1025 [Internet]. DOI: 10.3390/f10111025
- [214] Wilson CJ, William Carey J, Beeson PC, Gard MO, Lane LJ. A GIS-based hillslope erosion and sediment delivery model and its application in the Cerro Grande burn area. *Hydrological Processes*. 2001;**15**:2995-3010
- [215] Ruiz-Gallardo JR, Castaño S, Calera A. Application of remote sensing and GIS to locate priority intervention areas after wildland fires in Mediterranean systems: A case study from South-Eastern Spain. *International Journal of Wildland Fire*. 2004;**13**:241-252
- [216] Lewis SA, Wu JQ, Robichaud PR. Assessing burn severity and comparing soil water repellency, Hayman fire, Colorado. *Hydrological Processes*. 2006;**20**:1-16
- [217] Fox DM, Laaroussi Y, Malkinson LD, Maselli F, Andrieu J, Bottai L, et al. POSTFIRE: A model to map forest fire burn scar and estimate runoff and soil erosion risks. *Remote Sensing Applications*. 2016;**4**:83-91
- [218] Lanorte A, Cillis G, Calamita G, Nolè G, Pilogallo A, Tucci B, et al. Integrated approach of RUSLE, GIS and ESA Sentinel-2 satellite data for post-fire soil erosion assessment in Basilicata region (southern Italy). *Geomatics, Natural Hazards and Risk*. 2019;**10**:1563-1595
- [219] Meléndez-Pastor I, Pedreño JN, Lucas IG, Zorpas AA. A model for evaluating soil vulnerability to erosion using remote sensing data and a fuzzy logic system. In: *Modern Fuzzy Control Systems and its Applications*. London, UK: InTech; 2017
- [220] Pijl A, Reuter LEH, Quarella E, Vogel TA, Tarolli P. GIS-based soil erosion modelling under various steep-slope vineyard practices. *Catena (Amst)*. 2020;**193**:104604. DOI: 10.1016/J.CATENA.2020.104604
- [221] D'Oleire-Oltmanns S, Marzolf I, Peter KD, Ries JB. Unmanned aerial vehicle (UAV) for monitoring soil erosion in Morocco. *Remote Sensing*. 2012;**4**:3390-3416 [Internet] [cited 2023 Jan 11]. Available from: <https://www.mdpi.com/2072-4292/4/11/3390/htm>
- [222] Eltner A, Mulsow C, Maas H-G, Eltner A, Mulsow C, Maas H-G. Quantitative measurement of soil erosion from TIs and Uav data. *ISPAr*. 2013;**XL1**:119-124 [Internet] [cited 2023 Jan 11]. Available from: <https://ui.adsabs.harvard.edu/abs/2013ISPAr.XL1b.119E/abstract>
- [223] de Luis M, González-Hidalgo JC, Raventós J. Effects of fire and torrential rainfall on erosion in a Mediterranean gorse community. *Land Degradation and Development*. 2003;**14**:203-213
- [224] Connor SE, Araújo J, Boski T, Gomes A, Gomes SD, Leira M, et al. Drought, fire and grazing precursors to large-scale pine forest decline. *Diversity and Distributions*. 2021;**27**:1138-1151
- [225] Savadogo P, Sawadogo L, Tiveau D. Effects of grazing intensity

and prescribed fire on soil physical and hydrological properties and pasture yield in the savanna woodlands of Burkina Faso. *Agriculture, Ecosystems and Environment*. 2007;**118**:80-92

[226] Pausas JG, Ribeiro E, Vallejo R. Post-fire regeneration variability of *Pinus halepensis* in the eastern Iberian Peninsula. *Forest Ecology and Management*. 2004;**203**:251-259

[227] Pausas JG. Resprouting of *Quercus suber* in NE Spain after fire. *Journal of Vegetation Science*. 1997;**8**:703-706 [Internet] [cited 2023 Jan 12]. Available from: <https://onlinelibrary.wiley.com/doi/full/10.2307/3237375>

[228] Ne'eman G, Lahav H, Izhaki I. Recovery of vegetation in a natural East Mediterranean pine forest on Mount Carmel, Israel as affected by management strategies. *Forest Ecology and Management*. 1995;**75**:17-26

[229] Fernández-Guisuraga JM, Suárez-Seoane S, Calvo L. Modeling *Pinus pinaster* forest structure after a large wildfire using remote sensing data at high spatial resolution. *Forest Ecology and Management*. 2019;**446**:257-271

[230] Gitas I, Mitri G, Veraverbeke S, Polychronaki A, Gitas I, Mitri G, et al. Advances in remote sensing of post-fire vegetation recovery monitoring - A review. *Remote Sensing of Biomass - Principles and Applications*. 2012;**1**:334. Available from: <https://www.intechopen.com/state-item.id>

[231] Bullock EL, Woodcock CE, Olofsson P. Monitoring tropical forest degradation using spectral unmixing and Landsat time series analysis. *Remote Sensing of Environment*. 2020;**238**:110968

[232] Stueve KM, Cerney DL, Rochefort RM, Kurth LL. Post-fire tree

establishment patterns at the alpine treeline ecotone: Mount Rainier National Park, Washington, USA. *Journal of Vegetation Science*. 2009;**20**:107-120 [Internet] [cited 2023 Jan 12]. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1654-1103.2009.05437.x>

[233] Salvia M, Ceballos D, Grings F, Karszenbaum H, Kandus P. Post-fire effects in wetland environments: Landscape assessment of plant coverage and soil recovery in the paraná river delta marshes, Argentina. *Fire Ecology*. 2012;**8**:17-37 [Internet] [cited 2023 Jan 12]. Available from: <https://fireecology.springeropen.com/articles/10.4996/fireecology.0802017>

[234] Addink EA, van Coillie FMB, de Jong SM. Introduction to the GEOBIA 2010 special issue: From pixels to geographic objects in remote sensing image analysis. *International Journal of Applied Earth Observation and Geoinformation*. 2012;**15**:1-6

[235] Polychronaki A, Gitas IZ, Minchella A. Monitoring post-fire vegetation recovery in the Mediterranean using SPOT and ERS imagery. *International Journal of Wildland Fire*. 2013;**23**:631-642 [Internet] [cited 2023 Jan 12]. Available from: <https://www.publish.csiro.au/wf/WF12058>

[236] Mitri GH, Gitas IZ. Mapping post-fire forest regeneration and vegetation recovery using a combination of very high spatial resolution and hyperspectral satellite imagery. *International Journal of Applied Earth Observation and Geoinformation*. 2013;**20**:60-66

[237] Morresi D, Vitali A, Urbinati C, Garbarino M. Forest spectral recovery and regeneration dynamics in stand-replacing wildfires of central Apennines derived from Landsat time series. *Remote Sensing*. 2019;**11**:308 [Internet] [cited

2023 Jan 12]. Available from: <https://www.mdpi.com/2072-4292/11/3/308/htm>

[238] Villarreal ML, Norman LM, Buckley S, Wallace CSA, Coe MA. Multi-index time series monitoring of drought and fire effects on desert grasslands. *Remote Sensing of Environment*. 2016;**183**:186-197

[239] Pérez-Cabello F, Montorio R, Alves DB. Remote sensing techniques to assess post-fire vegetation recovery. *Current Opinion in Environmental Science & Health*. 2021;**21**:100251

[240] Veraverbeke S, Gitas I, Katagis T, Polychronaki A, Somers B, Goossens R. Assessing post-fire vegetation recovery using red–near infrared vegetation indices: Accounting for background and vegetation variability. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2012;**68**:28-39

[241] Lasaponara R, Abate N, Fattore C, Aromando A, Cardettini G, Di Fonzo M. On the use of Sentinel-2 NDVI time series and Google earth engine to detect land-use/land-cover changes in fire-affected areas. *Remote Sensing (MDPI AG)*. 2022;**14**:4723 [Internet]. DOI: 10.3390/rs14194723

[242] Rahman S, Chang HC, Magill C, Tomkins K, Hehir W. Spatio-temporal assessment of fire severity and vegetation recovery utilising Sentinel-2 imagery in New South Wales, Australia. *International Geoscience and Remote Sensing Symposium (IGARSS)*. 2019;**9960**-3

[243] Alegria C. Vegetation monitoring and post-fire recovery: A case study in the Centre inland of Portugal. *Sustainability (Switzerland)*. 2022;**14**

[244] Bright BC, Hudak AT, Kennedy RE, Braaten JD, Henareh KA. Examining post-fire vegetation recovery with

Landsat time series analysis in three western North American forest types. *Fire Ecology*. 2019;**15**

[245] Li X, Zhang H, Yang G, Ding Y, Zhao J. Post-fire vegetation succession and surface energy fluxes derived from remote sensing. *Remote Sensing (MDPI AG)*. 2018;**10**:1000 [Internet]. DOI: 10.3390/rs10071000

[246] Veraverbeke S, Somers B, Gitas I, Katagis T, Polychronaki A, Goossens R. Spectral mixture analysis to assess post-fire vegetation regeneration using Landsat thematic mapper imagery: Accounting for soil brightness variation. *International Journal of Applied Earth Observation and Geoinformation*. 2012;**14**:1-11

[247] Röder A, Hill J, Duguy B, Alloza JA, Vallejo R. Using long time series of Landsat data to monitor fire events and post-fire dynamics and identify driving factors. A case study in the Ayora region (eastern Spain). *Remote Sensing of Environment*. 2008;**112**:259-273

[248] Riaño D, Chuvieco E, Ustin S, Zomer R, Dennison P, Roberts D, et al. Assessment of vegetation regeneration after fire through multitemporal analysis of AVIRIS images in the Santa Monica Mountains. *Remote Sensing of Environment*. 2002;**79**:60-71

[249] Sankey TT, Moffet C, Weber K. Postfire recovery of sagebrush communities: Assessment using Spot-5 and very large-scale aerial imagery. *Rangeland Ecology & Management*. 2008;**61**:598-604

[250] Solans Vila JP, Barbosa P. Post-fire vegetation regrowth detection in the Deiva Marina region (Liguria-Italy) using Landsat TM and ETM+ data. *Ecological Modelling*. 2010;**221**:75-84

[251] Fernandez-Carrillo A, McCaw L, Tanase MA. Estimating prescribed fire

impacts and post-fire tree survival in eucalyptus forests of Western Australia with L-band SAR data. *Remote Sensing of Environment*. 2019;**224**:133-144

[252] Zhou Z, Liu L, Jiang L, Feng W, Samsonov SV. Using long-term SAR backscatter data to monitor post-fire vegetation recovery in tundra environment. *Remote Sensing (MDPI AG)*. 2019;**11**:2230 [Internet]. DOI: 10.3390/rs11192230

[253] Chhabra A, Rüdiger C, Yebra M, Jagdhuber T, Hilton J. RADAR-vegetation structural perpendicular index (R-VSPI) for the quantification of wildfire impact and post-fire vegetation recovery. *Remote Sensing (MDPI AG)*. 2022;**14**:3132 [Internet]. DOI: 10.3390/rs14133132

[254] Chen W, Jiang H, Moriya K, Sakai T, Cao C. Monitoring of post-fire forest regeneration under different restoration treatments based on ALOS/PALSAR data. *New Forest (Dordr)*. 2018;**49**:105-121

[255] Attarchi S, Gloaguen R. Classifying complex mountainous forests with L-band SAR and Landsat data integration: A comparison among different machine learning methods in the Hyrcanian Forest. *Remote Sensing*. 2014;**6**:3624-3647 [Internet] [cited 2023 Jan 13]. Available from: <https://www.mdpi.com/2072-4292/6/5/3624/htm>

[256] Carreiras JMB, Jones J, Lucas RM, Shimabukuro YE. Mapping major land cover types and retrieving the age of secondary forests in the Brazilian Amazon by combining single-date optical and radar remote sensing data. *Remote Sensing of Environment*. 2017;**194**:16-32

[257] Kuplich TM. Classifying regenerating forest stages in Amazônia using remotely sensed images and a neural network. *Forest Ecology and Management*. 2006;**234**:1-9

[258] Polychronaki A, Gitas IZ, Minchella A. Monitoring post-fire vegetation recovery in the Mediterranean using SPOT and ERS imagery. *International Journal of Wildland Fire*. 2014;**23**:631-642

[259] de Luca G, Silva JMN, Modica G. Short-term temporal and spatial analysis for post-fire vegetation regrowth characterization and mapping in a Mediterranean ecosystem using optical and SAR image time-series. *Geocarto International*. 2022;**37**(27):15428-15462. DOI: 10.1080/10106049.2022.2097482

[260] Lefsky MA, Cohen WB, Parker GG, Harding DJ. Lidar remote sensing for ecosystem studies: Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *Bioscience [Internet]*. 2002;**52**:19-30. DOI: 10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO

[261] Sato L, Gomes V, Shimabukuro Y, Keller M, Arai E, dos-Santos M, et al. Post-fire changes in forest biomass retrieved by airborne LiDAR in Amazonia. *Remote Sensing (MDPI AG)*. 2016;**8**:839 [Internet]. DOI: 10.3390/rs8100839

[262] Gordon CE, Price OF, Tasker EM. Mapping and exploring variation in post-fire vegetation recovery following mixed severity wildfire using airborne LiDAR. *Ecological Applications*. 2017;**27**:1618-1632 [Internet] [cited 2023 Jan 14]. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1002/eap.1555>

[263] Bolton DK, Coops NC, Wulder MA. Characterizing residual structure and forest recovery following high-severity fire in the western boreal of Canada using

Landsat time-series and airborne lidar data. *Remote Sensing of Environment*. 2015;**163**:48–60

[264] Meng R, Wu J, Zhao F, Cook BD, Hanavan RP, Serbin SP. Measuring short-term post-fire forest recovery across a burn severity gradient in a mixed pine-oak forest using multi-sensor remote sensing techniques. *Remote Sensing of Environment*. 2018;**210**:282–296

[265] Martín-Alcón S, Coll L, de Cáceres M, Guitart L, Cabré M, Just A, et al. Combining aerial LiDAR and multispectral imagery to assess postfire regeneration types in a Mediterranean forest. *Canadian Journal of Forest Research*. 2015;**45**:856–866

[266] Viana-Soto A, García M, Aguado I, Salas J. Assessing post-fire forest structure recovery by combining LiDAR data and Landsat time series in Mediterranean pine forests. *International Journal of Applied Earth Observation and Geoinformation*. 2022;**108**:102754. DOI: 10.1016/J.JAG.2022.102754

[267] de las Heras J, Moya D, Vega JA, Daskalakou E, Vallejo VR, Grigoriadis N, et al. Post-Fire Management of Serotinous Pine Forests. 2012 [cited 2023 Jan 14]. pp. 121–50. Available from: https://link.springer.com/chapter/10.1007/978-94-007-2208-8_6

[268] Pádua L, Guimarães N, Adão T, Sousa A, Peres E, Sousa JJ. Effectiveness of Sentinel-2 in multi-temporal post-fire monitoring when compared with UAV imagery. *ISPRS International Journal of Geo-Information*. 2020;**9**:225 [Internet] [cited 2023 Jan 14]. Available from: <https://www.mdpi.com/2220-9964/9/4/225/htm>

[269] Fernández-Guisuraga JM, Sanz-Ablanedo E, Suárez-Seoane S, Calvo L. Using unmanned aerial vehicles

in Postfire vegetation survey campaigns through large and heterogeneous areas: Opportunities and challenges. *Sensors*. 2018;**18**:586 [Internet] [cited 2023 Jan 12]. Available from: <https://www.mdpi.com/1424-8220/18/2/586/htm>

[270] Samiappan S, Hathcock L, Turnage G, McCraine C, Pitchford J, Moorhead R. Remote sensing of wildfire using a small unmanned aerial system: Post-fire mapping, vegetation recovery and damage analysis in grand bay, Mississippi/Alabama, USA. *Drones*. 2019;**3**:1–18

[271] Lazzeri G, Frodella W, Rossi G, Moretti S. Multitemporal mapping of post-fire land cover using multiplatform PRISMA hyperspectral and sentinel-UAV multispectral data: Insights from case studies in Portugal and Italy. *Sensors (MDPI AG)*. 2021;**21**:3982 [Internet]. DOI: 10.3390/s21123982

[272] Aicardi I, Garbarino M, Lingua A, Lingua E, Marzano R, Piras M. Monitoring post-fire forest recovery using multi-temporal digital surface models generated from different platforms. *EARSeL eProceedings*. 2016;**15**(1):1–8

[273] Talucci AC, Forbath E, Kropp H, Alexander HD, DeMarco J, Paulson AK, et al. Evaluating post-fire vegetation recovery in Cajander larch forests in Northeastern Siberia using UAV derived vegetation indices. *Remote Sensing*. 2020;**12**:2970 [Internet] [cited 2023 Jan 14]. Available from: <https://www.mdpi.com/2072-4292/12/18/2970/htm>

[274] Larrinaga AR, Brotons L. Greenness indices from a low-cost UAV imagery as tools for monitoring post-fire Forest recovery. *Drones*. 2019;**3**:6 [Internet] [cited 2023 Jan 14]. Available from: <https://www.mdpi.com/2504-446X/3/1/6/htm>

[275] Jain P, Coogan SCP, Subramanian SG, Crowley M, Taylor S, Flannigan MD. A

review of machine learning applications in wildfire science and management. *Environmental Reviews*. 2020;**28**:478-505

opportunities. *IEEE Geoscience and Remote Sensing Magazine*. 2022;**10**:270-294

[276] Chen J, Zheng W, Wu S, Liu C, Yan H. Fire monitoring algorithm and its application on the geo-Kompsat-2A geostationary meteorological satellite. *Remote Sensing*. 2022;**14**:2655 [Internet] [cited 2023 Jan 16]. Available from: <https://www.mdpi.com/2072-4292/14/11/2655/htm>

[277] Bailon-Ruiz R, Bit-Monnot A, Lacroix S. Real-time wildfire monitoring with a fleet of UAVs. *Robotics and Autonomous Systems*. 2022;**152**:104071. DOI: 10.1016/J.ROBOT.2022.104071

[278] Roteta E, Bastarrika A, Padilla M, Storm T, Chuvieco E. Development of a Sentinel-2 burned area algorithm: Generation of a small fire database for sub-Saharan Africa. *Remote Sensing of Environment*. 2019;**222**:1-17

[279] Crowley MA, Cardille JA. Remote Sensing's recent and future contributions to landscape ecology. *Current Landscape Ecology Reports*. 2020;**5**:45-57 [Internet] [cited 2023 Jan 17]. Available from: <https://link.springer.com/article/10.1007/s40823-020-00054-9>

[280] Chen Y-N, Fan K-C, Chang Y-L, Moriyama T. Special issue review: Artificial intelligence and machine learning applications in remote sensing. *Remote Sensing (MDPI AG)*. 2023;**15**:569 [Internet]. DOI: 10.3390/rs15030569

[281] Koubaa A, Ammar A, Abdelkader M, Alhabashi Y, Ghouti L. AERO: AI-enabled remote sensing observation with onboard edge computing in UAVs. *Remote Sensing (MDPI AG)*. 2023;**15**:1873 [Internet]. DOI: 10.3390/rs15071873

[282] Zhang L, Zhang L. Artificial intelligence for remote sensing data analysis: A review of challenges and

Section 2

Data Science in Geospatial Analysis and Applications

Spatial Cloud Computing Using Google Earth Engine and R Packages

*Anwarelsadat Eltayeb Elmahal and
Mohammed Mahmoud Ibrahim Musa*

Abstract

Google Earth Engine (GEE) is an open-source spatial cloud computing platform that can be easily accessed through its native application programming interfaces (APIs), which are JavaScript and Python. In this chapter, we will introduce the R package, a non-native GEE API, as a different way to interface with and utilize the extensive statistical and visualization capabilities of the R package through the RGEE library. In addition to the fundamental concepts of spatial cloud computing, the reader will be exposed to the practical side of satellite image processing techniques, which will include, digital image processing and visualizations.

Keywords: spatial cloud computing, Google Earth Engine, R package, RGEE library, satellite image processing, image classification, machine learning algorithms, accuracy assessment, NDVI

1. Introduction

Through the internet or other specialized networks, cloud computing has gained a lot of attention and is being increasingly utilized. It provides clients with applications, data, computing processing power, and information technology management as a service. Individuals could use it for emailing, data storage or photo sharing. It is used by enterprise businesses for a range of business operations as well as process management.

Geospatial data, particularly that obtained through remote sensing, is expanding at an exponential rate and using cloud computing services. For users to handle it professionally in the cloud, it requires both spatial and programming skills. Google Earth Engine (GEE) is a well-known example of a cloud-based geospatial platform that uses Google's multi-Petabyte data repository to let users analyze and visualize planetary-scale environmental data.

GEE could be easily accessed through its native application programming interfaces (APIs), which are JavaScript and Python. In this chapter, we will introduce the R package, a non-native GEE API, as a different way to interface with and utilize the extensive statistical and visualization capabilities of the R packages through the RGEE

library. In addition to the fundamental concepts of spatial cloud computing, the reader will be exposed to the theoretical and practical sides of remote sensing principles, and satellite image processing techniques, which will include, digital image processing and visualizations.

2. Principles and concepts

2.1 Remote sensing principles

We currently live in a world where the population is constantly growing, and natural resources are being continuously depleted by land use and landcover improper management. This situation leads to catastrophic climate change challenges that may threaten humanity's very existence. Understanding the Earth's system and applying that knowledge to inform our decisions can help us overcome these challenges [1, 2]. Remote sensing is an approach for collecting information about the ocean, land, and atmosphere utilizing electromagnetic radiation without directly contacting the object, surface, or phenomenon under investigation [3, 4]. The components of a satellite remote sensing system are (i) data collection via satellite platforms and sensors, (ii) data conversion to ground stations, (iii) data processing, and (iv) distribution to users after applying quality assurance standards (**Figure 1**).

After the official introduction of GEE in 2010 [5], simple methods for handling earth observation (EO) data became available. Contrary to the conventional methods of EO data manipulation, the new methods do not require the user to download or retain massive datasets, deal with a variety of file formats, or own highly developed computer systems in order to analyze and visualize EO data.

2.2 Analysis and visualization of satellite data

Images are an efficient way to communicate ideas, and numerous experts say that a picture is worth a 1000 words. Remotely sensed data are received in image format

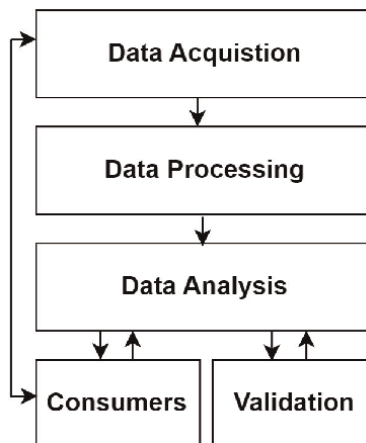


Figure 1.
Remote sensing components simplified.

and may contain some systematic or accidental errors as a result of atmospheric interaction. Before evaluating and interpreting remotely sensed data, the user must be aware of how to correctly pre-process it; otherwise, they face the risk of depending on images that have already been processed, unless they were handled by a professional. The satellite image processing techniques could be categorized into three main categories [6–8], which are:

- i. **Image restoration and rectification:** it refers to the operations that correct images from both geometric and radiometric errors. Image restoration may include georeferencing and haze reduction in satellite images. Image registration or restoration must be implemented to correct distorted or degraded image data. The unprocessed image data must first be processed in order to adjust for geometrical distortions, calibrate the data radiometrically, and eliminate noise [9].
- ii. **Image enhancement:** It means enhancing the image's quality so that it is easier for others to perceive it. This might be done by increasing the contrast in the image, using filtering, or using any other approaches.
- iii. **Information extraction and image classification:** The purpose of image classification is to convert the image's spectral classes into information classes [7, 8]. With the goal to categorize the pixels with the same spectral values to, for instance, known land use and land cover classes. There are typically two main categories, which are: (i) unsupervised and (ii) supervised. In the unsupervised category, the computer/machine will classify the spectral classes into information classes without any major user inputs. To get satisfactory results, the technique is followed by laborious post-classification operations. It is preferable when working with expansive areas or when the user has limited background information regarding the image's classes. However, with supervised classification, the user selects the training areas, and the computer/machine creates the classes based on those selections. In both categories, strict accuracy standards should be adhered by utilizing machine learning (ML) algorithms. In this chapter, a real example of image enhancement followed by classification is provided using the RGEE package.

2.3 Image classification and machine learning approaches

We employed a supervised image classification technique in the example at the end of this chapter, in which the training areas were designated for various land use land cover (LULC) classes. The location chosen for the test is Wadi Sob, a well-known watershed in Khartoum State, Sudan. The heart of ML is the exploration and creation of mathematical models and algorithms to learn from data and get insights out of that. In ML, the data is split into two subsets: a training test for building the model and a test subset to evaluate it. In this chapter, we chose the SmileCart algorithm because it generates clear findings that are simple enough to be comprehended, can be used to categorical and continuous datasets, is non-parametric, and makes no prior assumptions. The datasets were divided into two groups: 80% were used for training, and 20% were used to test and validate the algorithm's effectiveness.

2.4 Normalized Difference Vegetation Index (NDVI)

The normalized difference vegetation index (NDVI) is a biophysical measure. It measures the difference between the red light, which plants absorb, and near-infrared light, which plants reflect and is usually calculated by an empirical formula [7, 8].

$$NDVI = (NIR - R) / (NIR + R).$$

In this chapter, an example is provided to calculate the NDVI and visualize it using the rgee platform. The area selected for the test is Kassla State, Sudan.

3. Specificity of spatial data

3.1 Understanding the nature of spatial data

Geographical information science (GISc) is used to identify and characterize an area of science concerned with the analysis, visualization, and representation of geo-referenced data [2]. The terms spatial and geospatial data are interchangeably used to mean any data that contains information about a specific location implicitly at a global level or explicitly at the local levels up to the GPS coordinates.

Based on storage and representation approaches, spatial data can be categorized into two types: vector and raster data. Vector data represents spatial data as points, lines, and polygons, whereas raster data expresses it as pixels (**Figure 2**). The building block of the vector model is the point, a line is a combination of two points and a polygon is the closed loop of more than two lines. The building block of the raster model is the pixel (picture element). The pixel is the most tiny component of the digital display. The spatial resolution refers to the size of the pixel representing the geographic phenomenon. An image of 10 m resolution-like that of the Sentinel-2- meaning that every pixel is 10 m × 10 m width. Both vector and raster data contain attribute values. Raster data is mostly found in remotely sensed data in the form of satellite images, aerial photos, and digital elevation models.

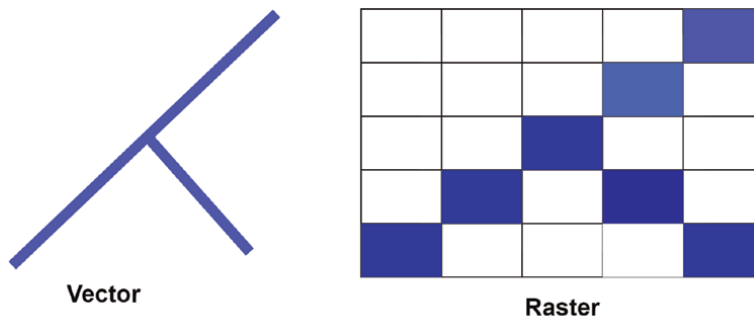


Figure 2. Vector (crisp/right) and raster (pixelated/left) models. (Source: https://pygis.io/docs/a_intro.html.)

3.2 Structure of spatial data

Spatial objects, specifically vectors are categorized as points, lines/poly-lines and Polygons, and as the building block is the point, once you read the new object you can view it by seeing the first couple of lines by applying these codes:

this code will allow you to see the first five lines of the data and having good idea about its attributes.

```
head(new object@ 'data sorucese, n = 5).
```

```
head(new object@polygons[[1]]@Polygons[[1]]@coords, 3).
```

4. Importance of R in data science and statistical analysis

4.1 R ecosystem

R is an open-source software and interpreted programming general-purpose language. R has expanded from a specific statistical language to a larger approach for data science and beyond the information technology and communication (ITC) industry [10]. The R programming language has emerged as the de facto data science programming language. Its adaptability, power, maturity, and expressiveness have made it an indispensable tool for data scientists worldwide [11]. Some of the several popular applications of R include, but are not limited to, ML; predictive modeling; data visualization; statistical modeling; scraping; dashboard construction; and interactive websites. Programmers could develop codes efficiently via an integrated development environment (IDE), R or Rstudio. The user will have a variety of R packages provided by the Comprehensive R Archive Network (CRAN) <https://cran.r-project.org/mirrors.html>.

Figure 3 shows some R essential packages used in data science.

Learning a strong programming language is crucial given the demand for data science, data engineering, and manipulation of big data. Python, R, SQL, and

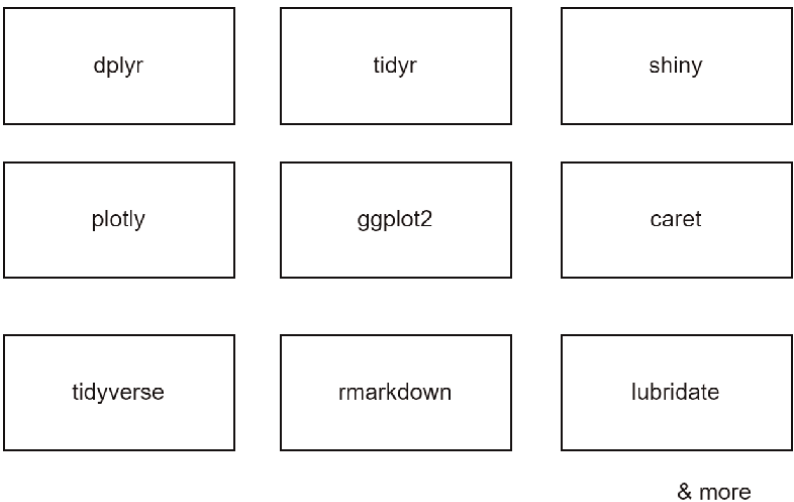


Figure 3.
Essential R packages for data science.

JavaScript are some of the most widely used programming languages for data research. R has not been as popular in recent years as Python, but it still attracts a lot of data scientists in academia since it excels at data wrangling, visualization, and statistical computing.

4.2 R spatial ecosystem

Spatial data conceptualization has been a part of R from its inception. The development of spatial packages that can handle and manipulate both raster and vector data is a continuous and progressive process [12]. For spatial data entry and analysis, numerous R packages have been released (**Figure 4**). `sp` and `sf`, among others, are used to manage basic features. For manipulating spatial data, `rgeos` and `rgdal` are the most common. The R-based Geospatial Data Abstraction Library (`rgdal`) is an extension of the Geospatial Data Abstraction Library (GDAL). For processing spatiotemporal data, the Spacetime library was established, and the Raster package was introduced for handling raster data.

4.3 Familiarity with spatial data in R

The user must have an R package, which could be downloaded from: <http://cran.r-project.org/>, make sure to use the most stable version, not necessarily the latest version. It is preferable to use an R editor like RStudio, or any alternative editor that you think will work for you. Posit Cloud provides a cloud-based environment for R (<https://login.rstudio.cloud>).

Despite the fact that there are thousands of packages, for the purpose of spatial data manipulation the user may start with these packages: `ggmap`; `rgdal`; `rgeos`; `tidyr`; `dplyr`; `maptool`; and `tmap`. To understand the functionalities of each package, use the help command in R or Rstudio to get the details. You can simply install all the packages by using the concatenate function and you can upload all the packages as well by using the `lapply` function in R.

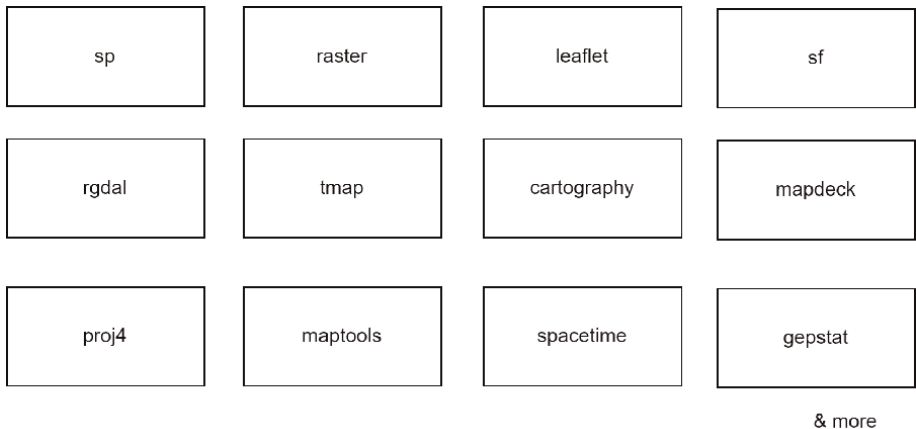


Figure 4.
Essential R packages for spatial data analysis.

```
# Assign a variable and concatenate the packages to install them.  
x <- c('rgdal','rgeos','maptools','ggmap','tidyr','dplyr','tmap','ggmap').  
# load all the needed packages once.  
lapply(x,library,character.only = TRUE).
```

4.4 Manipulating spatial data with R

One of the most important features of R is the ability to handle spatial data, for instance reading shapefiles. The library (rgdal), as mentioned previously, as the R extension for the GDAL library will allow users to manipulate spatial data, you could read it by using readOGR function as per the example in the code below:

```
# loading the rgdal library for manipulating spatial data.  
library(rgdal).  
new object <- readOGR(dsn = 'data source name', layer = 'the layer').
```

The first line of code is for calling the rgdal library, while the second line of code is used to load the shapefile while assigned to a new object named “new object” in which the readOGR function is the function to load the file with both the data source name and the layer names identified. You can try it using your own data. Make sure to set up the working directory for R properly to read the data without complications. The head function is usually used to display the first lines, the default output is the first six lines, but in the above code are only the first three lines as identified by the number 3.

5. Cloud computing

Cloud computing attracted users and enterprises as it simplifies ITC services because storage and operating systems do not need to be installed or maintained locally, a third party is taking care of that at a low cost. Cloud computing refers to the on-demand, pay-as-you-go availability of computer system resources, particularly data storage and processing power, without direct active control by the user [13, 14]. In Cloud computing, there are three service models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) [14, 15]. IaaS is regarded as the foundation of the cloud service, PaaS is the platform for constructing applications, and SaaS is where the providers manage every part of the products and clients are only concerned with accessing the software at their convenience.

5.1 Spatial cloud computing

Spatial cloud computing is a cloud computing paradigm based on geospatial sciences and spatiotemporal principles [16]. There are four service models for spatial cloud computing [17], which are: (1) Data as a Service (Daas), (2) Model as a Service (Maas), (3) Geoprocessing as a Service (Gaas), and (4) Workflow as a Service (WaaS). The focus of this chapter is considered as PaaS, as it serves as a platform for developers to create and run scalable applications on the Google Cloud platform.

6. Google Earth Engine platform

GEE is a cloud-based platform to analyze planetary-scale environmental data, that utilizes Google's multi-Petabyte data catalog [18], GEE is categorized as PaaS and used

by scientists, researchers, and developers to identify changes, plot trends, and measure variability on planetary surfaces. The GEE platform is an all-inclusive development and deployment environment for spatial cloud computing that falls under the PaaS category. The main components of GEE are: (i) Datasets provided by different agencies, mainly NASA, USGS and the European Space Agency; (ii) Computer power provided by Google Inc.; (iii) REST APIs/client libraries for making request to the EE; and (iv) The code editor as an IDE for developers using JS API.

6.1 GEE data catalog

Earth engine as illustrated in **Figure 5**, is composed of an open data catalog, computational infrastructure, geospatial APIs and an interactive application server [5]. Utilizing the petabytes of spatial data archives provided by US government agencies, the European Space Agency, and many others is now possible as a result of the introduction of GEE as a spatial cloud computing platform. Below are some descriptions of the available datasets provided by GEE.

6.2 Landsat data

The Landsat Program is the most extensive and ongoing space-based record of Earth's land in existence [19]. The Landsat satellites records offer broad spectral data extending from visible light to short-wave infrared, as well as covering a long time period (1972-present). The majority of the data's spatial resolution is in the 30 m range, although it can vary depending on the Landsat series and bands that are chosen. Landsat data is ideal for analyzing changes in land use and land cover as it offers both rich data and a large time span for tracking changes on the surface of the globe.

6.3 MODIS data

MODIS (or Moderate Resolution Imaging Spectroradiometer), Terra and Aqua satellites view the entire earth every 1–2 days [20]. The first MODIS was launched in 1999 and acquires the data in 36 bands, with spatial resolution ranges from 250 to 1000 m and spectral bands ranging from 0.4 to 14.4 μm . MODIS excels at observing significant biosphere changes ("MODIS Web," n.d.). It is perfect for meteorological and climate change activities.

6.4 Sentinel data

The Sentinel was launched in 2014. The Sentinel missions carry different sensors such as radar and multi-spectral imaging instruments for land, ocean and atmospheric

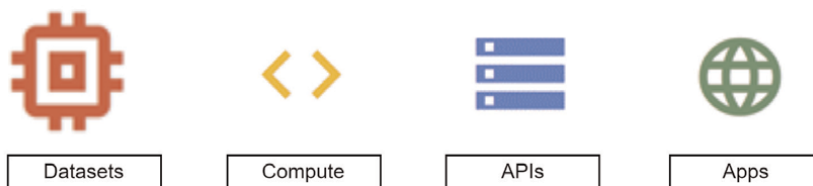


Figure 5.
Earth engine components. (Source: earthengine.google.)

monitoring [21]. The sensors are used for depicting land surface changes, inventory of land resources and depicting pollution on land, water and atmosphere.

7. Interacting with Google Earth Engine

The user needs to interact with GEE via specific APIs to download, process, analyze, and visualize Google Earth Catalog data. There are two native APIs that users usually use to interact with GEE which are JavaScript (JS) and Python APIs. In the following sections, we will explore the `rgee` package, an Earth Engine client library for R [22], which enables R users with specific functions to call GEE through the Python API.

API stands for API, it is a contract that would allow two or more computer programs to communicate with each other [23] or simply integrate different software systems [24]. GEE could be accessed easily via two native APIs, which are the JavaScript and the Python APIs. In this chapter, we will dig deep in exploring the potentiality of `rgee` package that integrates R and GEE via an intermediate package called `Reticulate`. You can think of a `reticulate` package as a bridge that connects R and GEE via the Python API.

JavaScript is one of the native GEE APIs and is an API that enables GIS developers to interact with the GEE platform via the Code Editor <http://code.earthengine.google.com>. The code editor is the web-based IDE for EE JS. The requirement to access the code editor is to register for Earth Engine <https://earthengine.google.com/> and to have at least an intermediate familiarity and experience with JavaScript coding. There are many learning platforms that beginners can start with in learning JS programming, but <https://www.w3schools.com/js/default.asp> and <https://developers.google.com/earth-engine/guides/getstarted> are good websites to start with.

Similarly, Python API is the other native GEE API that enables GIS developers to interact with the GEE platform, but via `Earth Engine(ee)` package and Google Collab platform (<https://colab.research.google.com/notebooks/>) or Jupyter Notebook (<https://jupyter.org/>).

To start the communication with GEE you need to import the `ee` library and trigger the authentication workflow `ee.Authenticate()`, and initialize the `ee` library `ee.Initialize()`. There are many learning websites, but a good start could be the one provided by Google: <https://developers.google.com/earth-engine/tutorials/community/intro-to-python-api>.

7.1 RGEE package bridging Earth Engine and Python API

The `rgee` package is a non-native R client library that employs the `reticulate` package of tools for Python-R interoperability [24] as cited in [18]. The `rgee` package enables users to combine the advantages of the R spatial ecosystem and GEE in a single workflow [18]. **Figure 6** explains how the R ecosystem could be integrated with GEE via the `reticulate` package and making use of google cloud utilities and assets.

8. Getting access to Google Earth Engine

To get access to GEE, sign up the form in: <https://signup.earthengine.google.com>, you will receive an explanatory email on how to get started. Or you have another

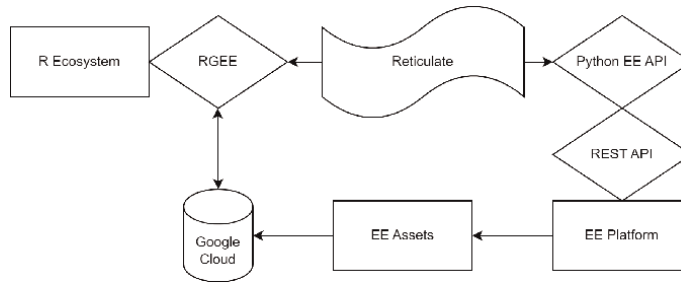
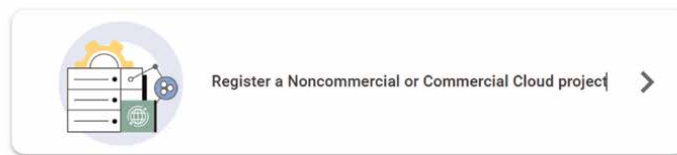


Figure 6.
The integration between the R ecosystem and GEE via the reticulate package.

Get started using Earth Engine

Earth Engine, Google's geospatial science platform in Google Cloud, is available for **paid commercial use** and **remains free for academic and research use**. Learn more about [Google Cloud projects](#).

Let's get started:



Noncommercial users can also use Earth Engine without creating Cloud projects.
[Click here for the signup form.](#)

Have an existing project? [Click here to go to the Code Editor](#)

Figure 7.
A screenshot of the GEE welcoming message.

option to try the code editor, once opening the code editor will receive a welcoming message (**Figure 7**), sign up and complete the registration process if needed.

8.1 Trying the code editor

The GEE code editor is a web-based IDE. **Figure 8** shows the components of the code editor and for more information and practice you can visit: <https://developers.google.com/earth-engine/guides/playground>

The code editor has many features (**Figure 7**) that can help the user to take advantage of the GEE API. The user can examine, display, create, manage, and exchange data and scripts.

8.2 rgee installation and initialization

The RGEE package's main purpose is to leverage the workflow between the R ecosystem and GEE, but it provides other functionalities [18] such as input/output design, interactive map display, time series extraction and asset management

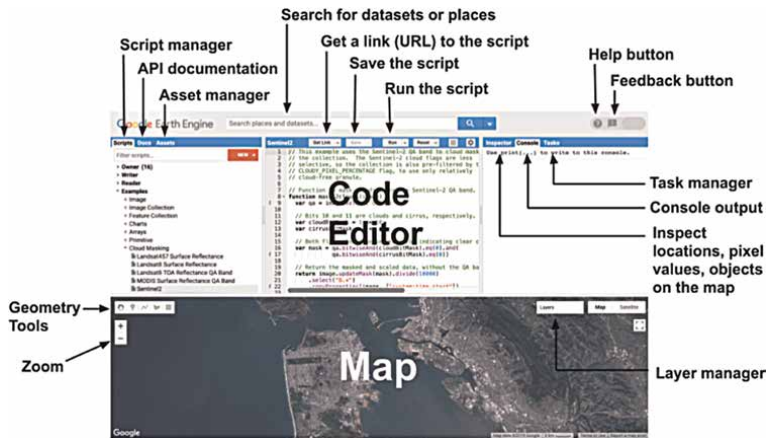


Figure 8.
 The components of the JavaScript code editor. (Source: <https://developers.google.com/earth-engine>.)

interface. To run and install RGEE properly Python environment version 3.5 or higher with NumPy and Earth Engine-API packages are required, you can download the Python version that suits your system from: <https://www.python.org/downloads/>, for more details about) Earth Engine Python API visit: https://developers.google.com/earth-engine/guides/python_install. In addition to python environment the user should have an active Google account, if you do not have one, you can create it from: www.google.com and the user should have access to Earth Engine platform and datasets, either via free non-commercial or commercial access, for more details visit: <https://developers.google.com/earth-engine/guides/access>. To use Earth Engine-API in the local environment, users should download and install Google Cloud SDK, which can be downloaded from: <https://cloud.google.com/sdk/docs/install>. To bridge the Python API with the R software, the user must install and upload the reticulate package, followed by installing and uploading remotes packages and finally, installing and uploading the RGEE package which could be downloaded from the GitHub repository (<https://github.com/r-spatial/rgee>). There are many ways by which you can install and upload the above-mentioned packages, but installation codes are a straightforward way of doing so. Different codes are provided to install the rgee package from the specified GitHub repository. The first step before rgee installation is to make sure that the reticulate package and Python 3.5 or greater are installed. Install the most stable version, not necessarily the latest version, as the stable versions are more mature. It is recommended that you install Python and create your own working environment to avoid any interoperability issues with other running applications. Below are the codes for installing the reticulate package and making sure that Python is installed and testing the environment. Once the reticulate package and python are installed you can make some simple operations using some Python libraries, like numpy, scipy, etc.

```
# install the reticulate package if it is not installed yet.
if (!(("reticulate" %in% installed.packages()[,1])) {
  print ("Installing package `reticulate`...")
  install.packages("reticulate")
}
```

```

else {
  print("The reticulate package is already installed")
}
library(reticulate).
# Testing python envirnment.
Sys.which("python").
# check if Python 3 installed.
Sys.which("python3").
# Making sure the python popular libraries are functional with reticulate.
np <- reticulate::import("numpy", convert = FALSE).
# do some array manipulations with NumPy.
a <- np$array(c(1:4)).
print(a).

```

In the above code, the first three lines instruct the program to install the ‘reticulate’ package if not installed yet. The package enables R to extend Earth Engine Python API classes and functions, you could use the library() function to load the reticulate package in your R eco-system. The next couple of lines load the ‘reticulate’ package to the system and test the installed Python package. You can import as many Python packages to your system according to your needs with the same previous code that you used in importing the NumPy package and you can just substitute the name of the packages using the reticulate function. Below are the codes to import the sklearn package.

```

reticulate::import("sklearn").
# Same function could be used to import other libraries from Python.

```

The remotes package functionality is to allow installation from remote repositories (below code). A confirmation message will be received, you may need to update some needed packages. Once installed and updated, cross-check the installed packages and make sure, the reticulate, R6 and processx packages are all installed, if not, do install them. Once all is done perfectly and after executing the initialization function, R will ask you to enter the verification code (**Figure 9**). Then a window will open in your web browser to add your GEE credentials details in the Notebook authenticator. Once credentials are provided, click on generate a token. In the R environment add generated token as verification code.

```

# Install Remotes Packages.
install.packages("remotes").
# Load Remotes Package into your R eco-system.
library("remotes").
# Install and Load RGEE packages.
remotes::install_github("r-spatial/rgee").
library(rgee).
# Install RGEE dependencies packages.

```



Figure 9.
Initialization of GEE.

```
ee_install().  
# Make Sure all dependencies installed properly.  
ee_check().  
# Authorize and Initialize RGEE API.  
ee_initialize().
```

The first two lines of the code instruct the application to load and install remote packages which will enable you to install RGEE package from GitHub. Once installed it should be uploaded to the system. The `ee_install()` function from the RGEE package would be used to install the RGEE dependencies and the `ee_check()` function is used to check that all the RGEE dependencies are installed. Once all is installed and loaded properly, a message will be received to inform you about the Python installations with the available packages installed. The final line of the code above, instructs the system to initialize the RGEE package which will bridge the gap between the R-ecosystem and the Python API to facilitate the process of interacting with GEE using R statistical and visualization power.

9. Case studies

In this section, we are going to walk through three case studies. The first case study is how to visualize Sentinel-2 datasets provided by GEE. It is a straightforward code that will allow the user to read, upload and visualize the Sentinel-2 data.

The second case study is how to apply the Normalized Vegetation Difference Index (NDVI) for the legendary Landsat Data. The NDVI is a commonly used measurement to determine the density and health of plants from remotely sensed data. It is determined using spectrometric data of the red and infrared regions of the electromagnetic radiations. We selected Kassala state in Sudan as it has a history of dry spells and is one of the drought-prone regions in Eastern Sudan.

The third case study is about the implementation of supervised classification of LULC for a large watershed in Sudan, Wadi Soba. The user may need to be aware that the supervised learning needs an appropriate selection of the training areas needed for the classification process, this may need expert knowledge or a thorough desk review to interpret the classes. We also applied the SimpleCart ML algorithm as it provides easily interpretable and visualizable results. The study area of the Soba watershed was chosen as the area witnessed urban expansions in the last decades that may negatively impact the area in the future.

9.1 Case study 1: Calculating NDVI for Sentinel-2 images

This case study will showcase how to use RGEE in uploading and visualizing the Sentinel-2 data (**Figure 10**). The `ee$ImageCollection` method is used to add image collection from the GEE repository. The `ee$FeatureCollection` is used to load vector data, which could be applied to clip the region of interest (ROI), which is part of Kassala State, Sudan in our case, it is an administrative boundary. The following code illustrates the process of uploading the sentinel image for Kassala State in Sudan using the RGEE library and GEE. But the user can substitute the data we used with the data that suits his/her purpose.

```
# load the RGEE API.  
library(rgee).  
# 2 Initialize RGEE API.
```



Figure 10.
Adding a Sentinel-2 image using the *rgee* package for Kassala State, Eastern Sudan.

```
ee_initialize().
# 3 Load study area Kassala admin unit will be used as study area.
study_area <- ee$FeatureCollection("FAO/GAUL/2015/level2")$filter(ee$Filter
$eq("ADM2_NAME", "Kassala")).
# 4 Load Sentinel 2 Image.
sentinel2_msi <- ee$ImageCollection("COPERNICUS/S2")$filterBounds
(study_area)$mean().
# 5 Define visualization parameters in an object literal.
vizParams <- list(
bands = c("B5", "B4", "B3"),
min = 0, max = 3000).
# 6 set study area as center of Map.
Map$centerObject(sentinel2_msi$clip(study_area)).
# 7 Add Sentinel Image to the Map.
Map$addLayer(sentinel2_msi$clip(study_area), vizParams, "S2").
```

The RGEЕ package in the above code is loaded into the R ecosystem in the first line using the `library()` function, and it is initialized in the second line using the `ee_initialize()` function. Line 3 uses the `filter()` function to select the Kassala area in eastern Sudan as the ROI while the `ee$FeatureCollection()` function loads level 2 administrative boundaries from the Earth Engine repository. In line 4, the `ee$ImageCollection()` class is used to load the sentinel 2 image, the `filterBounds()` function is used to identify image boundaries, while the `mean()` function is used to add mean value from available images in the specified study area. In line 5, object literal is used to define visualization parameters, `bands` property is used to select certain bands from the image, and `min` and `max` functions are used to specify the minimum and maximum values of pixels. In line 6, the `centerObject()` function is used to adjust the image to be at the center of the scene. Finally, in line 7, the `addLayer()` function is used to add an image to the scene.

In the second illustration, the NDVI calculation procedure is shown for a region in Kassala State, Sudan (**Figure 11**). The NDVI uses band math by correlating the red and infrared wavelengths of electromagnetic radiation to assess the healthiness of the vegetation cover in an area. The code below shows case how the NDVI could be calculated using GEE and RGEE interface.

```
# 1 load the R GEE API.
library(rgee).
# 2 Initialize RGEE API.
ee_Initialize().
# 3 Load Landsat 8 Image.
lc_image <- ee$Image("LANDSAT/LC08/C02/T1/LC08_171049_20220910").
# 4 setting visualization parameters.
vizParams <- list(
  min = 0,
  max = 0.5,
  bands = c("B5", "B4", "B3").
)
# 5 Setting map center and Zoom Level.
Map$setCenter(36.1899, 15.5010, 10).
# 6 Add Image to the Map.
Map$addLayer(lc_image, vizParams).
# 7 Simple Single-Band Vizualization.
ndvi <- lc_image$normalizedDifference(c("B4", "B5")).
# 8 Setting Visualization parameters.
ndviViz <- list(
  min = 0.5,max = 1,
  palette = c("00FFFF", "0000FF")).
# 9 Masking.
ndviMasked <- ndvi$updateMask(ndvi$gte(0.4)).
# 10 Create Region of Interest Geometry.
roi <- ee$Geometry$Point(c(36.1899, 15.5010))$buffer(20000).
# 11 Set Map center to region of intersets.
Map$centerObject(lc_image$clip(roi)).
# 12 Using region of interest to Clip and Add Landsat image to the Map.
Map$addLayer(
  eeObject = lc_image$clip(roi),
  visParams = vizParams,
  name = "Landsat 8".
)
# 15 Using region of interest to Clip and Add NDVI to the Map.
Map$addLayer(
  eeObject = ndviMasked$clip(roi),
  visParams = ndviViz,
  name = "NDVI"
)
```

The above code explains, how RGEE package is uploaded to the R-ecosystem in the first line, while the third line explains the way the Landsat 8 image is uploaded. Lines 4 and 5 explain the process of adding the visualization parameters and how to center the image to the ROI in this case the circle selected from Kassala State. In line 7, the `normalizedDifference()` function is used to calculate NDVI, Band 4, represents the red

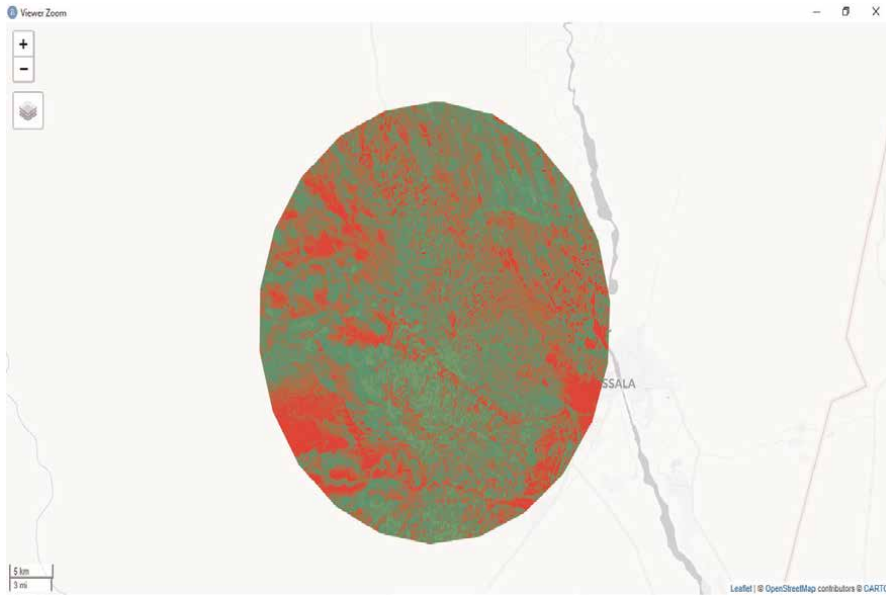


Figure 11.
Adding the NDVI layer to the map.as of September 2022.

wavelength from the visible light and Band 5, represents the infra-red portion of the electromagnetic radiation. Lines 8 and 9 indicate the codes for visualizing, masking and displaying NDVIs values more than 0.4, while buffered to the ROI in line 10. The rest of the code explains the adjustments to center the map and zooming to the ROI. The final portion of the code, below is to show how to add the NDVI layer to GEE interface using the RGEE package.

```
# 16 Using region of interest to Clip and Add NDVI Mask to the Map.
Map$addLayer(
  eeObject = ndviMasked$clip(roi),
  visParams = ndviViz,
  name = "NDVI"
)
```

10. Supervised classification

The principle behind supervised classification is that a user can choose sample pixels from an image (training areas) that are representative of particular classes. These sample pixels can then be used to find comparable pixels with the same spectral signatures and classify them into similar classes. To ensure that the categorization process produces the desired results, the training regions must be carefully chosen. Since the process of choosing the training region is widely described in the literature on remote sensing, we did not concentrate too much on it; rather, the focus of this chapter is on how we could run the algorithm using the RGEE package.

In addition, the reader will learn how to utilize the SmileCart Algorithm as a form of ML through the provided code below. The code has been implemented in the Wadi Soba Watershed of Khartoum State, Sudan. However, it can also be applied by the

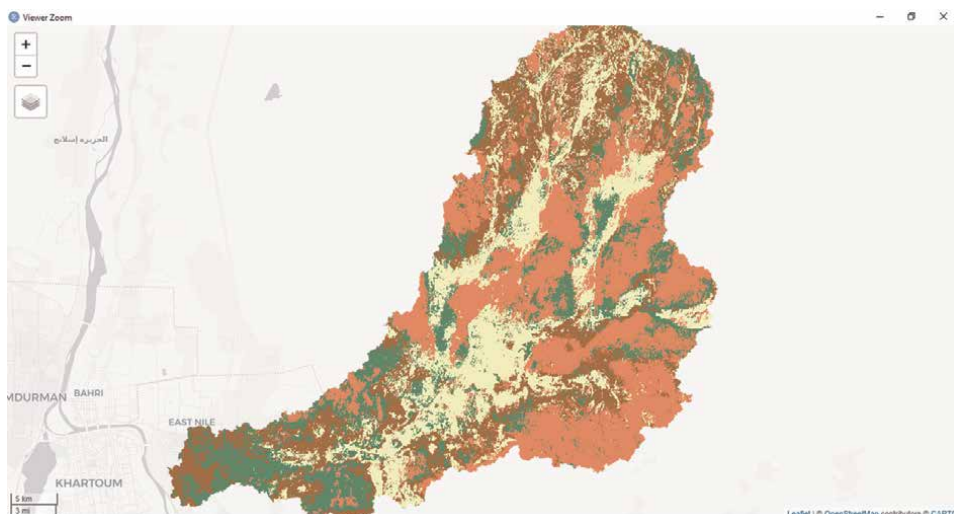


Figure 12.
 Showing the LULC classes for Wadi Soba.

reader in other areas. The first steps for adding images, visualizing them and clipping them to the desired regions of interest are nearly identical to the previous codes. When it comes to ML, dividing data into training and testing sets is key. The training set is used to predict classes, while the testing set is used to validate results and determine overall classification accuracy. The output of the code is shown in **Figure 12**.

```
# Load GREE.
library(rgee).
# Initialize RGEE.
ee_initialize().
# Add Region of Interest to the project and # You may need to store your datasets
in your own EE assets.
roi <- ee$FeatureCollection('projects/ee-anwareltayeb2/assets/
WADISOBA_WS').
# Add Classes and # You may need to store your datasets in your own EE assets.
Water <- ee$FeatureCollection('projects/ee-mohammedaau/assets/soba_water').
Agric <- ee$FeatureCollection('projects/ee-mohammedaau/assets/soba_agric').
Urban <- ee$FeatureCollection('projects/ee-mohammedaau/assets/soba_urban').
Barren <- ee$FeatureCollection('projects/ee-mohammedaau/assets/
soba_burren').
Rangeland <- ee$FeatureCollection('projects/ee-mohammedaau/assets/
soba_rangeland').
# Add Image.
image <- ee$ImageCollection("LANDSAT/LT05/C02/T1_TOA")$filterDate
('1984-2001-01', '1984-2012-31')$filter(ee$Filter$bounds(roi))$sort
('CLOUD_COVER')$first().
Soba84 <- image$select(c('B4', 'B3', 'B2')).
Soba84vis <- list(min = 0.0, max = 0.4, gamma = 1.2).
Map$addLayer(image$clip(roi), Soba84vis, 'SobaTM84False').
Map$centerObject(roi, 10).
label <- 'Class'.
```

```
bands <- c('B1','B2','B3','B4','B5','B6','B7').
input <- image$select(bands).
training <- Water$merge(Agric)$merge(Urban)$merge(Barren)$merge(Rangeland).
trainImage <- input$sampleRegions(training).
trainingData <- trainImage$randomColumn().
trainSet <- trainingData$filter(ee$Filter$lessThan('random',0.8)).
testSet <- trainingData$filter(ee$Filter$greaterThanOrEquals('random',0.8)).
classifier <- ee$Classifier$smileCart()$train(trainSet,label,bands).
classified <- input$classify(classifier).
landcoverPalette <- c(
  "#2c7bbc", # Water.
  "#4dac26", # Agric.
  "#d7191c", # Urban.
  "#ffffbc", # Barren.
  "#b8e186" # Rangeland).
Map$addLayer(classified$clip(roi),list(palette = landcoverPalette,min = 0,
max = 5),"classification84").
```

11. Conclusion

GEE, opened a new era in satellite image processing, analysis and visualization, as it is an open-source spatial cloud computing platform, which is freely available for users and researchers to access its resources via its native APIs, which are JavaScript and Python APIs. In this chapter, the rgee package was introduced as a different way to interface and utilize the extensive statistical components, and visualization capabilities of R software. The reader will be introduced to the Normalized Vegetation Difference Index (NDVI), one of the most biophysical indices in monitoring vegetation cover and plants' health over time. In addition to the fundamentals of spatial cloud computing, the reader will also be exposed to the practical side of installing, initializing and running GEE under R. A very good illustration of supervised classification utilizing the rgee package and ML methods is also provided.

Author details

Anwarelsadat Eltayeb Elmahal¹ and Mohammed Mahmoud Ibrahim Musa^{2*}

1 Faculty of Geography and Environmental Sciences, University of Khartoum, Sudan

2 Faculty of Computer Science and Information Technology, Alzaiem Alazhari University, Sudan

*Address all correspondence to: mohammed.aau@gmail.com

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Liang S, Li X, Wang J. Advanced Remote Sensing: Terrestrial Information Extraction and Applications. Cambridge, MA, USA: Academic Press; 2012
- [2] Longley P. Geographic Information Systems and Science. 2nd ed. Hoboken, NJ, USA: John Wiley & Sons; 2005
- [3] Fotheringham AS, Rogerson PA. The SAGE Handbook of Spatial Analysis. London, UK: SAGE Publications Ltd.; 2008
- [4] Martin S. An Introduction to Ocean Remote Sensing. Cambridge, U.K: Cambridge University Press; 2004
- [5] Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment Journal*. 2017; **202**:18-27. ISSN: 0034-4257/2017
- [6] Weng Q. An Introduction to Contemporary Remote Sensing. New York, USA: McGraw Hill Professional; 2011
- [7] Campbell JB, Wynne RH. Introduction to Remote Sensing. 5th ed. New York, USA: Guilford Press; 2011
- [8] Aronoff S. Remote Sensing for GIS Managers. Redlands, California, USA: ESRI Press; 2005
- [9] Chandra AM. Higher Surveying. Delhi, India: New Age International; 2005
- [10] Peng RD. R Programming for Data Science. Lulu.com. 2016
- [11] Bivand RS. Progress in the R ecosystem for representing and handling spatial data. *Journal of Geographical Systems*. 2021;**23**:515-546. DOI: 10.1007/s10109-020-00336-0
- [12] Ren K. Learning R Programming. Birmingham – Mumbai: Packt Publishing Ltd.; 2016
- [13] Huang Q. Spatial cloud computing. In: Wilson JP, editor. The Geographic Information Science & Technology Body of Knowledge. 2nd Quarter 2020 ed. New York, USA; 2020. DOI: 10.22224/gistbok/2020.2.7
- [14] Marinescu DC. Cloud Computing: Theory and Practice. Waltham, MA, USA: Elsevier; 2013
- [15] Yang C, Huang Q. Spatial Cloud Computing: A Practical Approach. Boca Raton, Florida, USA: CRC Press; 2013
- [16] Yang C, Goodchild M, Huang Q, Nebert D, Raskin R, Xu Y, et al. Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing? *International Journal of Digital Earth*. 2011;**4**:305-329
- [17] Jaatun MG, Zhao G, Rong C. Cloud computing. In: First International Conference, CloudCom 2009, Beijing, China, December 1-4, 2009, Proceedings, vol. 5931. Springer, 2009
- [18] Aybar et al. rgee: An R package for interacting with Google Earth Engine. *Journal of Open Source Software* 2020;**5**(51):2272. DOI: 10.21105/joss.02272
- [19] Landsat Science [WWW Document], n.d. Available from: <https://landsat.gsfc.nasa.gov/> [Accessed: 05 April 2023]
- [20] MODIS Web [WWW Document], n.d. Available from: <https://modis.gsfc.>

nasa.gov/about/ [Accessed: 05 April 2023]

[21] The Sentinel missions [WWW Document], n.d. Available from: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/The_Sentinel_missions [Accessed: 05 April 2023]

[22] Jacobson D, Brail G, Woods D. APIs: A Strategy Guide. Sebastopol, CA, USA: O'Reilly Media, Inc.; 2012

[23] Biehl M. API Architecture. Zürich, Switzerland: Create Space Independent Publishing Platform; API-University Press; 2015

[24] Ushey K, Allaire J, Tang Y. reticulate: Interface to 'python'. 2020. Retrieved from: <https://CRAN.R-project.org/package=reticulate>

A Probabilistic Approach for Road Network Analysis

*David Barilla, Giuseppe Caristi, Roberto Guarneri
and Dario Lo Bosco*

Abstract

Complex networks are networks whose structure is irregular, complex and which evolves over time and are used in various branches of science and technology, such as in biochemistry, in the study of interactions in quantum field theory, in the study of IT processes, topologies in geographical databases and also on the web, in social networks such as Facebook and Linkedin and in the Google model. The term “complex” itself derives from the Latin cum (together)—plexus (intertwined), “intertwined together”: it highlights that a complex network system is composed of a set of parts connected and “intertwined” in such a way that the result (the effect produced) is different from the sum of the constituent parts. Therefore, the behavior of a complex system cannot be inferred by a simple analysis of the elements that compose it, but it is necessary to carry out a systematic examination of the interactions that are generated between them and the constraints that determine their operation. In this chapter we show how the probability of intersections for the Road Network Analysis (RNA) can be useful. We use a geometric probabilities approach for transportation planning operations. We show the utility of the probability of intersections in the determination of a classification rule for raster conversions in Geographical Information System (GIS) and GRASS GIS.

Keywords: probabilistic approach, geometric probabilities, road network analysis, graph network, Geographical Information System (GIS)

1. Introduction

The realization of a public investment in mobility and logistic networks can be tackled by solving a problem of maximizing the “global utilities” associated with the planned intervention, through special algorithms that allow the systematic evaluation of the set of variability characteristics and constraints, expressed by equations and inequalities.

Constraints can also express conditions of nonnegativity of variables, such as when they represent physical quantities, or upper and lower limits on the value they can assume in the range of variation (budget, resources, etc.) [1–6].

These constraints can be algebraically represented by the conditions to which the action variables are subject and can be expressed by means of equations or inequalities, classifiable as:

- *Sign constraints* used to indicate that the variables can assume positive value or equal to 0

$$x \geq 0 \quad \text{and} \quad y \geq 0 \quad (1)$$

- *Operational constraints* used to indicate the maximum availability of resources, also expressed by equations or inequalities.

This approach must also consider the many probabilistic variables obtained from risk analysis in the various phases of operation and management, up to the decommissioning of the work, for the entire reference time scenario, in the interval $[0, T]$, with $T = \text{useful life years}$ of investment plans (such as the more complex ones in the PNRR).

The variables of the mathematical model used to optimize the results of the problem represent the unknowns on which the decision must be made and are characterized as “decision variables”, represented by specific indicator.

$$r_{ij}, \text{con} \begin{cases} i = 1, \dots, n; \\ j = 1, \dots, m \end{cases} \quad (2)$$

The objective function representative of the problem can be minimized, as in the case of the cost flow analysis C_i relating to the network infrastructure in question and the negative “impacts” on the ecosystem, or maximized if it is revenue flows or material and/or intangible benefits (including the so-called positive “externalities”, i.e. positive changes in utility) B_j , which must then be “discounted” in the economic calculation to outline the decision-making framework from an LCCA perspective—Life Cycle Cost Analysis (**Figure 1**) [7, 8].

Therefore, having evaluated the flows of global benefits and those of global costs, both discounted, the calculation of the cost-benefit ratio can be expressed with the relationship:

$$B_c = \frac{\sum b_{g_i}^{-w_i} \left(\frac{1}{(1+r)^t} \right)}{\sum c_{g_j}^{-w_j} \left(\frac{1}{(1+r)^t} \right)} \pm V_r \left(\frac{1}{(1+r)^N} \right), \quad (3)$$

With

$$\sum w_i = 1 \quad \text{and} \quad \sum w_j = 1. \quad (4)$$

The term $V_r \left(\frac{1}{(1+r)^N} \right)$ of the expression indicates the residual value of the network infrastructure that is the subject of the investment, at the end of its useful life, estimated equal to “N” years. This can take on positive values, as in the case of the sale

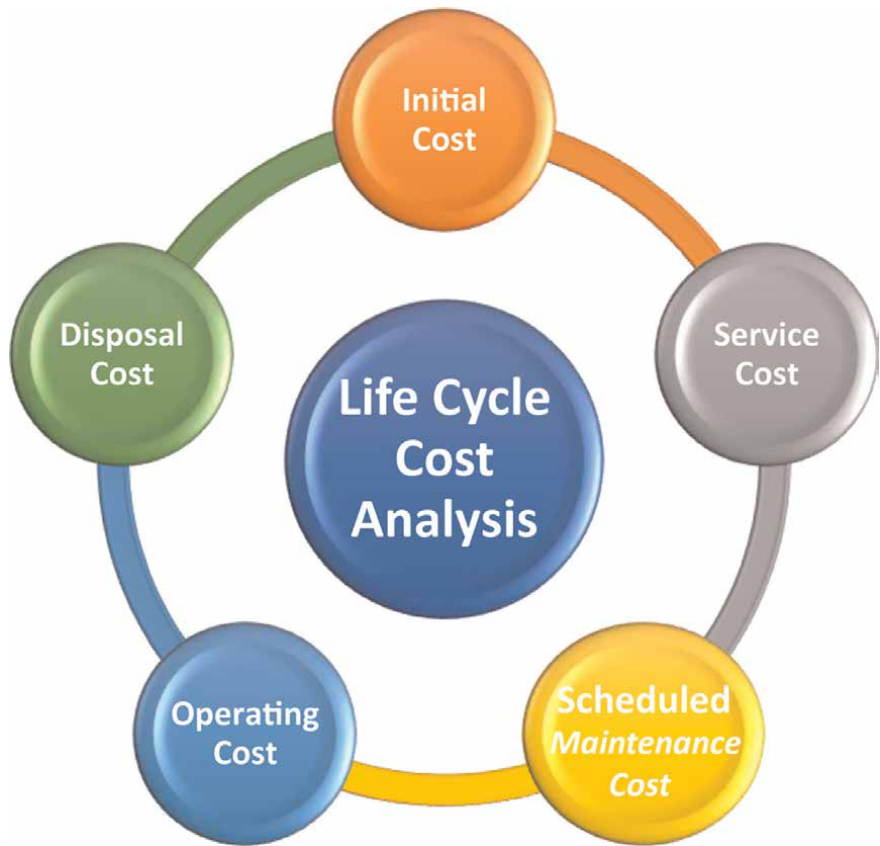


Figure 1.
LCCA—Life Cycle Cost Analysis.

and/or reuse of the asset in a production cycle, or negative, when a cost must be borne for disposal, possible rehabilitation of the area where the public work is located, etc.

Indicating, then, with C_0 the discounted initial investment cost if any $(B_c - C_0) > 1$ the investment will be economically convenient and can be profitably implemented [8, 9]. For an n -variable mathematical model to maximize the objective function $\vec{f}$ it will be necessary to solve the typical maximum search problem

$$\max \left[\vec{f}(x_1, x_2, \dots, x_j, \dots, x_n) \right]. \quad (5)$$

In the systematic analysis to be conducted for investments in mobility and logistics networks, since the security of the “network system” is paramount, among the characteristic variables it is necessary to accurately evaluate all the probabilistic indicators of security.

$$P_X^s(x) \quad (6)$$

which can be defined algebraically by means of the general mathematical relation:

$$P_X^s(x) = f_X \left(C_{gr}, L_{tr}, G_C^{i=1, \dots, n}, A_r^{i=1, \dots, n}, G_S^{i=1, \dots, k}, T_r^{i=1, \dots, n} \right). \quad (7)$$

This formula expresses the relationship between global risk management costs C_{gr} , the level of tolerability of the same L_{tr} , the degree of connection of the branches of the mobility network under study $G_C^{i=1, \dots, k}$ and the relative accessibility of said branches $A_r^{i=1, \dots, n}$, as well as the degree of intrinsic safety $G_S^{i=1, \dots, k}$ of the k constituent elements of the infrastructure in question and the contribution made to safety by the technological devices operating in the network n itineraries that make up the transport system in question, which are indicated by $T_r^{i=1, \dots, n}$.

Furthermore, the indicator $A_r^{i=1, \dots, n}$ used above must be mathematically “weighted”, by introducing into the expression described a special factor “ ε ” associated with the specific intrinsic conditions of grid reliability, for the management of emergencies in critical situations (significant weather-climatic events, earthquakes, landslides, structural failure of works, etc.).

Finally, in the case of non-homogeneity of the level of risk tolerability L_{tr} in the road axes and nodes of the network, the variable L_{tr} will in turn be a function of the different specific levels of risk tolerability, according to the following relationship which also considers the degree of interconnection of the nodes of the graph in which the network is geometrically represented (Figure 2).

Denoting with β_{r_i} the characteristic variable for the generic path “network path” which connects node n_i by axes r_i considered in the mobility system.

$$\beta_{r_i} = \sum \left(\beta_{i \rightarrow j}^{i=1, \dots, n} \right)^{-w_{\beta_i}} \quad (8)$$

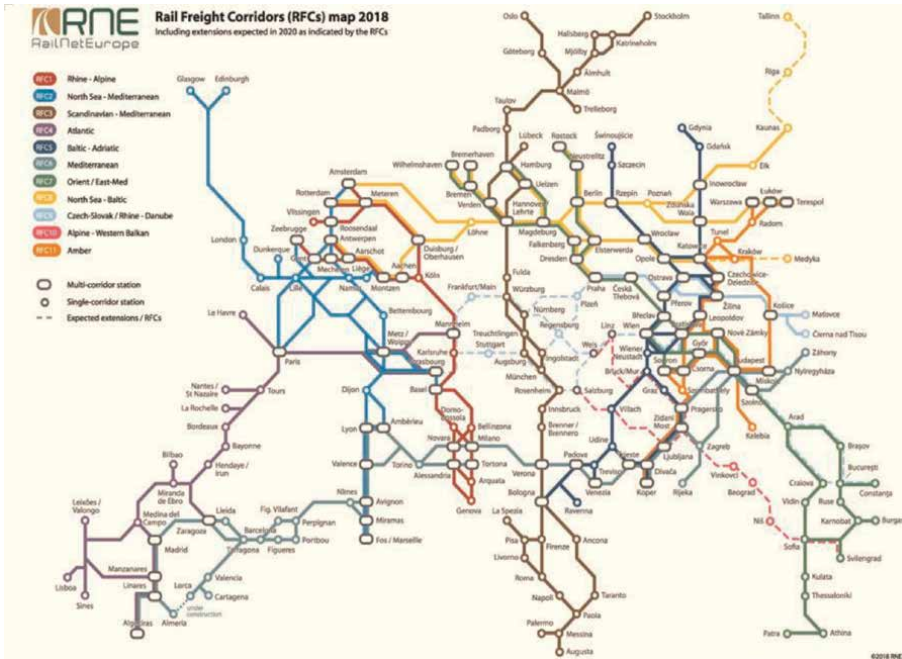


Figure 2.
Graph of EU freight corridors established by regulation EU/913/2010.

According to this approach, the risk tolerance level defined above L_{tr} is expressed by the following relationship:

$$L_{tr} = \left[g \left(L_{tr_i}^{r,n} \right) \right] \beta_{ri}. \quad (9)$$

2. Graph network geometric model

On the basis of the mathematical formulation of the problem defined above, for the best representation of the problem, it is useful to carry out an appropriate modeling of the infrastructural network through a special “graph” characterizing the various its design components, such as the O-D paths (origin-destination), network (arcs of the graph) and “centroids” generating and/or attracting traffic (nodes of the graph), as in **Figure 2** [10].

Indeed, this procedure makes it possible to carry out a systematic examination of the many variables characterizing the problem under study, through a relational structure composed of a finite set of objects (on a more strictly mathematical level, a finite set of points) and of a set relationship between pairs of them.

In particular, if the pair of nodes is ordered, i.e. if the arcs have an arrival node and a departure node, the graph is said to be “directed” and an ordered sequence of arcs identifies the “paths”.

If a quantitative characteristic is also associated with the latter, the graph gives rise to a particular mobility or logistics “network”. By operating, for example, a specific mathematical characterization for the railway network, it is possible to carry out a study of accessibility, even in an emergency (following relevant events), using two specific indicators:

- Single Accessibility Index (IAS)_i;
- Combined Accessibility Index (IAC)_i.

The variable (IAS)_i represents the accessibility value of each arc “i” of the railway network, without considering the interactions between contiguous arcs (the accessibility value of neighboring arcs is not considered).

It may happen that an arc has a limited or zero accessibility value but is interposed between two highly accessible arcs: it will then be characterized by an accessibility value (even null), regardless of whether the two neighboring arcs are accessible and allow therefore to also access the arc in question.

The combined accessibility index (IAC)_i, on the other hand, considers the mutual interactions between contiguous arcs.

In this way, within the model, the possibility is considered that an arc with limited accessibility can still be reached using the neighboring (contiguous) arcs with greater accessibility, evaluated through the coefficient k .

The representation of the network with special evaluated graphs, integrated by particular mathematical models of stochastic geometry and geometric probabilities [11–15], such as those that will be described below, is an important decision aid tool for investment analysis or for solving design, management and maintenance problems, and above all for the optimization of operating safety, even in critical conditions, due to accidental obstacles along the network path, due to adverse meteorological events (landslides, instability, falling rocks, etc.), or seismic events, etc.

Methodologically, the problem can be analytically tackled through an appropriate interpretation of the “graph” constructed with “attributes” and thought of as formed

by a whole union of elementary regular and/or irregular geometric figures, constituting special lattices R , in the geometric space S^m .

By means of specific test bodies (mathematical models) representative of the vehicles (passenger and/or freight trains for railway networks, or cars, busses, articulated lorries or other types of vehicles for road networks, etc.), it is possible to study the relative motion given speed V in the lattice R and study any “interferences” produced by “obstacles” (of given shape and size) that may be found along the generic arc $h-k$ of the geometric lattice associated with the mobility network under examination.

These test bodies can be assumed as segments of suitable length l (to schematize a convoy with a large number of carriages, as for example occurs in the composition of a freight train, etc., or of an articulated lorry in the case of road transport), or made up of rectangles with sides l_1 and l_2 (as in the case of a high-speed train, a regional train, etc., or even a car along a street).

To arrive at a mathematical resolution of the problem, we will assume in the following developments that each side of the considered lattice offers the same resistance capacity to the advancement of the test-body (model).

Furthermore, to operationally address the problems in professional practice, reference can also be made in the schematization of the “lattices R union of elementary forms” to a “catalog” illustrated below, built on the results of scientific



Figure 3 - The high-speed train network in East Asia



Figure 3.
The high-speed train network in East Asia.



Figure 4.
Tree on the roadway.



Figure 5.
Accidental obstacles on a railway site.

studies carried out on the different forms of the regular and irregular geometric figures constituting a generic network, such as the rail network in **Figure 3**, referring to East Asian high-speed trains.

To then define the accidental obstacles on site (railway or road) and interfering with the movement of vehicles, a different generic shape is assumed (for example, triangular or circular like the section of a tree trunk accidentally fallen across a road), or a multi-sided polygon, representative of other possible cases¹ (**Figures 4 and 5**).

The mathematical criterion described above makes use of a regular lattice formed by a union of equal regular polygons obtained starting from a figure called the “fundamental cell of the lattice”, with only segments of the geometric boundaries in common.

If in the vertices of the fundamental cell we consider some regular equal polygons, called “obstacles”, we then obtain a “regular lattice with obstacles”.

As a basic example for the interpretation of the method we can consider the fundamental cell formed by a square with side A and, as an obstacle, a square with side a with $2a < A$ (**Figure 6**)

If we randomly throw a segment s of constant length l onto the lattice, the problem consists in determining the probability that such a segment s intersects one side of the lattice. This probability is equal to the probability that a segment s intersects the boundary of the fundamental cell C_0 (**Figure 7**).

We have that $\text{area } C_0 = \frac{c^2}{2}$ and we consider the limit positions of s , with $\widehat{C_0(\rho)}$ being the figure determined by such positions (**Figure 8**).

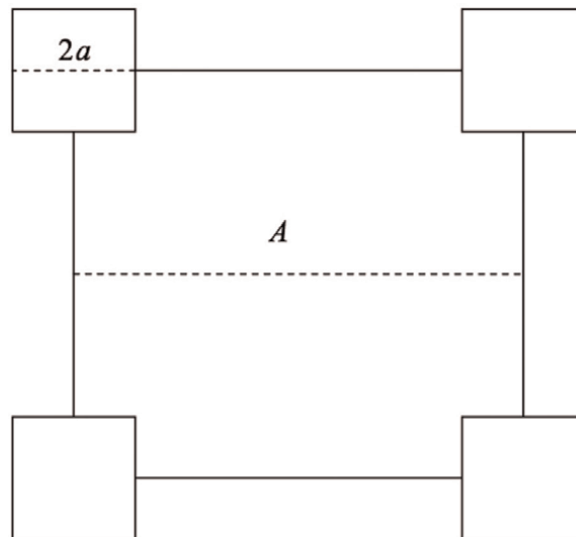


Figure 6.
Lattice with square obstacles.

¹ For the purposes of operating safety, a road case is very different from a “guided guide” railway case: indeed, as soon as the motorist perceives the presence of an obstacle along the road, he can try to carry out a maneuver to by-pass it (where the site and traffic conditions allow it), while the driver of a train has the only possibility of trying to stop the movement of the train before its impact on the track on which it travels.

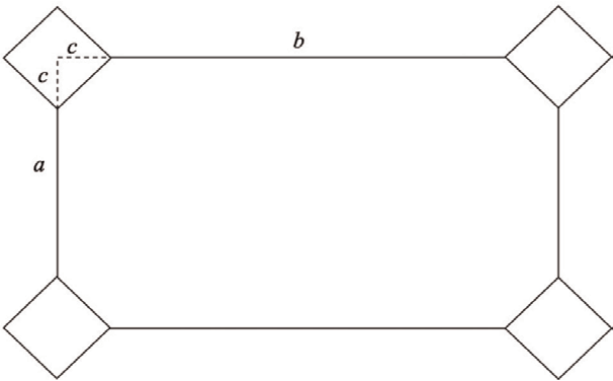


Figure 7.
 Lattice with rhombus obstacles.

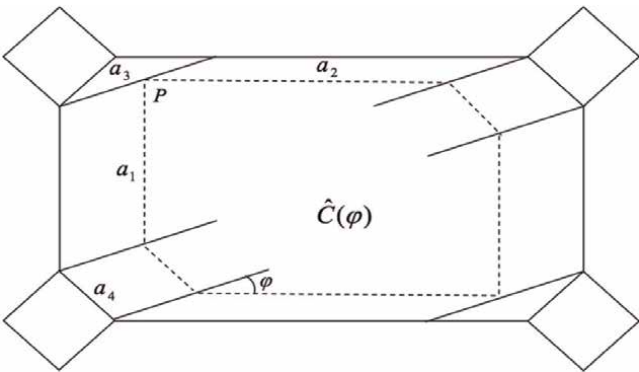


Figure 8.
 Lattice of limit positions.

Computing the areas, we obtain

$$area\widehat{C_0}(\rho) = ab - \frac{c^2}{2} - \left[(a - c)l\cos\rho + \left(b - \frac{c}{2}\right)l\sin\rho + \frac{l^2}{2}\sin 2\rho \right]. \quad (10)$$

Denoting with M_1 the set of segments s which have their middle point in C_0 and with N_1 the set of segments s which have their middle point in C_0 but do not intersect the frontiers of C_0 . Then, the probability is

$$P_{int} = 1 - \frac{\mu(N_1)}{\mu(M_1)}, \quad (11)$$

where μ is the *Lebesgue* Measure in Euclidean plane [16].
 The measures $\mu(M_1)$ and $\mu(N_1)$ are computed by the *Poincaré* Kinematic measure [17].

$$dk = dx \wedge dy \wedge d\rho, \quad (12)$$

where x, y are the coordinates of P and φ the defined angle.

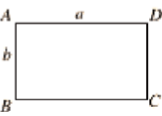
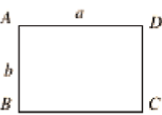
Since $\varphi \in [0, \frac{\pi}{2}]$, by previous formula we have that:

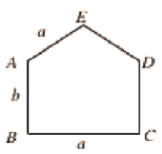
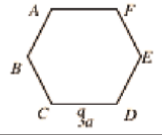
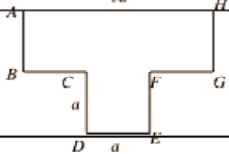
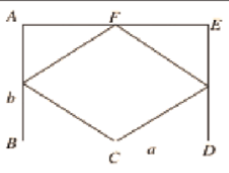
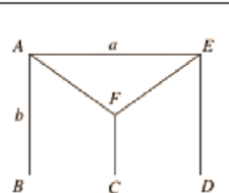
$$P_{int} = \frac{(a + b - \frac{c}{2})l - l^2}{\pi(ab - \frac{c^2}{2})} \quad (13)$$

In place of the fundamental cell and the obstacles (mathematical models) considered above, we can consider others representative of the numerous typical cases that it is possible to examine in the case of complex networks, both road and railway, or other material and/or immaterial networks (such as web networks, social networks, economic-financial networks, etc.).

For the practical applications of the method, the following “catalog” allows us to be able to consider even irregular grids, given that the conformation of the grid can be formed by a whole union of various geometric figures of all shapes and sizes.

The segment s considered (“test body”) in the problem of geometric probabilities can represent—as mentioned—a railway freight train, or a wheeled articulated lorry, but in other cases other test bodies can be considered, for example rectangles, such as the one that represents a passenger train (including those for HS/HC) or, in the road case, a simple car (two-dimensional figure, since one dimension is not negligible compared to the other) (Table 1).

Elements Catalogue of the road and/or railway network graph and geometric probability			
Network element	Obstacle	Body test	Geometric Probability
Regular Polygons			
		—	$P = \frac{2(2a + 2b)l - l^2}{2\pi ab}$
			$P = \frac{2[(n+1)a + b]l - l^2}{\pi(n+1)ab}$
	  	—	$P = \frac{2(2a + b)l - l^2}{2\pi ab}$

			$P = \frac{2(a+2b)l - l^2}{2\pi ab}$
			$P = \frac{2 \left\{ l \left[2a \cos^2 \alpha + b(1 - \sin \alpha) \right] - \frac{\cos \alpha (2 \cos \alpha - \sin \alpha)}{4} l^2 \right\}}{\left(\frac{\pi - \alpha}{2} \right) a \cos \alpha (a \sin \alpha + 2b)}$
			$P_{in}^{(1)} = \frac{2(4a+2b)l - l^2}{8\pi ab}$
			$P_{in}^{(2)} = \frac{2 \left(4a + 2b - \frac{3m}{2} \right) l - \frac{3l^2}{2} - \frac{3\pi(\pi-2)m^2}{16}}{\pi \left(6ab - \frac{3m^2}{4} - \frac{3\pi m^2}{16} \right)}$
			$P_{in} = \frac{1}{2\alpha \left[2ab - m^2 \left(1 + \sin 2\alpha + \cos \alpha + \cos \frac{\alpha}{2} \right) \right]} \left\{ 2 \left[a(1 - \cos \alpha) + 3b \sin \alpha \right] l - 3lm(1 - \cos \alpha) - \frac{l^2}{2} \right. \\ \left. \left[(3 + 2 \sin \alpha) \cos^2 \alpha + \left(\frac{b}{a} + \cos \alpha - \frac{1}{2 \cos \alpha} \right) \sin 2\alpha + \frac{\alpha}{\cos \alpha} \right] - \alpha n \right\}$
			$P_{in} = \frac{1}{\alpha \left[4ab - 2m^2 \left(\sin 2\alpha + \cos \alpha + \cos \frac{\alpha}{2} + \frac{\pi}{2} \right) \right]} \left\{ 2 \left[a(1 - \cos \alpha) + \right. \right. \\ \left. \left[\left(4 \cos \alpha + \sin \frac{\alpha}{2} \right) (1 - \cos \alpha) + \sin \alpha \left(1 + \cos \frac{\alpha}{2} - 4 \sin \alpha \right) \right] l \right. \\ \left. \left. - \frac{l^2}{2} \left[(3 + 2 \sin \alpha) \cos^2 \alpha + \left(\frac{b}{a} + \cos \alpha - \frac{1}{2 \cos \alpha} \right) \sin 2\alpha + \frac{\alpha}{\cos \alpha} \right] \right\} \right. \\ \left. \frac{1}{\frac{\alpha ab}{2} - m^2 \alpha (1 + \sin \alpha + \cos \alpha)} \left\{ -al \sin \frac{\alpha}{4} \sin \frac{3\alpha}{4} \left(3 \cos \frac{3\alpha}{4} + ct, \right. \right. \right. \\ \left. 4bl \sin \frac{\alpha}{4} \sin \frac{3\alpha}{4} + lm \left(1 - \cos \frac{\alpha}{2} + 2 \sin \frac{\alpha}{4} \sin \frac{3\alpha}{4} \right) - \frac{l^2}{4} \left[4 + tg \frac{\alpha}{2} \right. \right. \right. \\ \left. \left. \left. (2 \cos \alpha + tg \alpha + ctg \alpha) \sin \frac{\alpha}{2} \cos \frac{3\alpha}{2} + \frac{\alpha}{2} (2 \cos \alpha + tg \alpha + ctg \alpha \cdot \right. \right. \right. \\ \left. \left. \left. \frac{\alpha}{4} (1 + \sin \alpha + \cos \alpha) m^2 \right) \right\} \right\}$

			$P_{\text{car}} = \frac{1}{\frac{a^3}{2} \left(6\alpha - \frac{5\pi}{4} \right) \tan \alpha - \frac{m^3}{4} \left[2\alpha (\sin \alpha + \cos \alpha + 2 \sin 2\alpha) - \frac{\pi}{2} \cos \alpha \right]}$ $\left\{ l \left[\frac{a}{\cos \alpha} (\cos \alpha - \cos 2\alpha - 3 \sin^2 \alpha \cos 2\alpha) - m \sin^3 \alpha \right] - \frac{l^2}{4} \left[1 + 4 \cos^2 2\alpha - \tan \alpha \cos 2\alpha \left(2\alpha - \frac{\pi}{2} - \sin 2\alpha \right) - (4\alpha - \pi) \cot 2\alpha \right] - \frac{m^2}{4} (4\alpha - \pi) \sin 2\alpha \right\}.$
			$P_{\text{car}}^{(1)} = \frac{1}{\alpha \left[(a + b \cos \alpha) \cdot b \sin \alpha - \frac{m^2}{2} \sin \alpha \right]}$ $\left\{ [b \sin^2 \alpha + (a + 2b \cos \alpha)(1 - \cos \alpha)] l - (1 - \cos 2\alpha) \frac{l^2}{4} - \frac{m^2}{4} \alpha \sin \alpha \right\}.$
			$P_{\text{car}} = 1 - \frac{\mu(\mathcal{N})}{\mu(\mathcal{M})},$ $\mu(\mathcal{M}) = \int_0^a d\varphi \iint_{\{x,y \in C_0\}} dx dy = \int_0^a (\text{area } C_0) d\varphi = \alpha \text{area } C_0,$ $\mu(\mathcal{N}) = \int_0^a d\varphi \iint_{\{x,y \in N_0(\varphi)\}} dx dy = \int_0^a [\text{area } \hat{C}_0(\varphi)] d\varphi = \text{area } C_0 - [a \sin \alpha$ $\frac{m}{2 \sin \beta} \{ a \sin \alpha [\sin(\alpha + \beta) - \sin 2\alpha + \sin \alpha + \sin \beta] +$ $b [(1 + \sin \alpha - \cos \alpha) \sin \beta - (\sin \alpha + \sin \beta \cos \alpha) \cdot \sin(\alpha +$ $+ \frac{a^2}{8 \sin \beta} \left[2 \cos^2 \alpha (2 \sin \alpha + \cos \frac{\alpha - \beta}{2}) - 2 \alpha \cot \alpha (\sin \alpha + \sin \beta - 2 \cos$ $+ \frac{ab}{8 \sin \beta} \left[2 (\cos \alpha + \cos \beta) + 4 \cos \frac{\alpha - \beta}{2} (\cot \alpha + \cot \beta) \right]$ $\sin 2\alpha + (\sin \alpha + \sin \beta) (\cos 2\alpha - 1) + \alpha (\sin \alpha + 18 \sin \beta$ $+ \frac{b^2}{4} (\cos 2\alpha + \cot \beta \sin 2\alpha + 2 \alpha \cot \beta - 1).$
			$P_{\text{car}} = \frac{1}{\left(\frac{\pi}{2} - 2\alpha \right) ab} - 2 \left(a - \frac{b \tan \alpha}{2} \right) (\cos \alpha - \sin \alpha) - 2 (\cos \alpha - \sin \alpha$ $\frac{1}{4} [3l^2 + (4 + \tan \alpha)m^2 - (\tan \alpha + \cot \alpha)lm] \cos 2\alpha -$ $\frac{1}{4} \left(\frac{\pi}{2} - 2\alpha \right) [(\tan \alpha + 2 \cot \alpha)l^2 + (4 \tan \alpha + 2 \cot \alpha)m^2 - 4lm].$

Table 1.
Geometric probabilities applied to network security for different means of transport.

3. Extension model to immaterial networks and social network analysis

The mathematical model described above for material road and railway networks can be extended to applications for immaterial networks, such as Social Network Analysis (SNA) and to the study of the so-called “mental map”, for problem analysis and as a tool decision aid [18–22].

Indeed, any graphical representation of the thought and theories of the Social Sciences. Always places the accent on a particular dimension of social reality, making use of a reticular structure formed by nodes and arcs that connect them, giving rise to a specific graph of the immaterial network.

Social networks, as a representation of a reticular society or resulting from constraints and opportunities emerging from the relationships between subjects, give rise to graphs of complex networks,² such as the one shown in the following figure, where the individual subjects from which the various input/output solicitations originate/arrive are represented at the nodes, while the interrelationships are expressed by the various “branches”, following a mathematical approach scheme like that of an “O-D” (origin-destination) matrix of a transport network.

The content generated by the users of social networks has given birth to a new era of Business Intelligence & Analytics (BI & A) centered on textual analysis and unstructured content. SNA today increasingly represents a great opportunity for companies to interact bilaterally with the market, much more effective than traditional “one-way” marketings and “business-to-market” type techniques.

Furthermore, both in the economic and sociological fields and for applied sciences in general, the typological structuring of so-called “mental maps”, consisting of the graphic translation of the paths taken by the human brain for the elaboration of a thought, i.e. “conceptual maps”, in the exposition of a problem or a complex and articulated concept with interconnected logical frameworks, also lends itself well to the modeling interpretation provided here for making decisions.

Through the mind mapping technique, it is possible to describe the complex process in which the mind associate's ideas in a non-linear way, processes information in a creative, intuitive, and figurative way, giving rise to a typical map with a radial structure, in the form of a particular oriented graph [23].

Some arcs are directed towards the most important node, represented by the thought processing center (the man who “travels” with his mind): the logical reference scheme is of the “input-output” type, with complex mechanisms of analytical reiteration (**Figure 9**).

In the network there may also be associative relationships between the various nodes situated on the same level, but the previous and subsequent ones are given by precise structured hierarchical links (**Figure 10**).

These tools allow you to put meta-reasoning into practice (allowing you to reflect on thought in an “introspective” way) and have practical applications that can be used not only to study, to memorize, or to plan (project management, etc.), but are of scientific and professional interest in various fields of application, such as forensic technology, engineering, economic and environmental sciences, artificial intelligence, medicine (treatment of dyslexia, dysgraphia, learning disabilities) and today they also find profitable use in the world of the web.

² The research sectors interested in the network optimization approach proposed here are, among others, Networks and Social Services, Networks and Social Capital, Networks and Social Control, Networks and Migrations, Networks and Health, Networks and Work, Networks and Education, Networks and Crime, Networks and Communication, Internet Networks and Virtual Communities, Networks, Organization and Management, Networks and Power, Scientific Research Networks and Technological Innovation, Corporate, Economic and Financial Networks, etc.

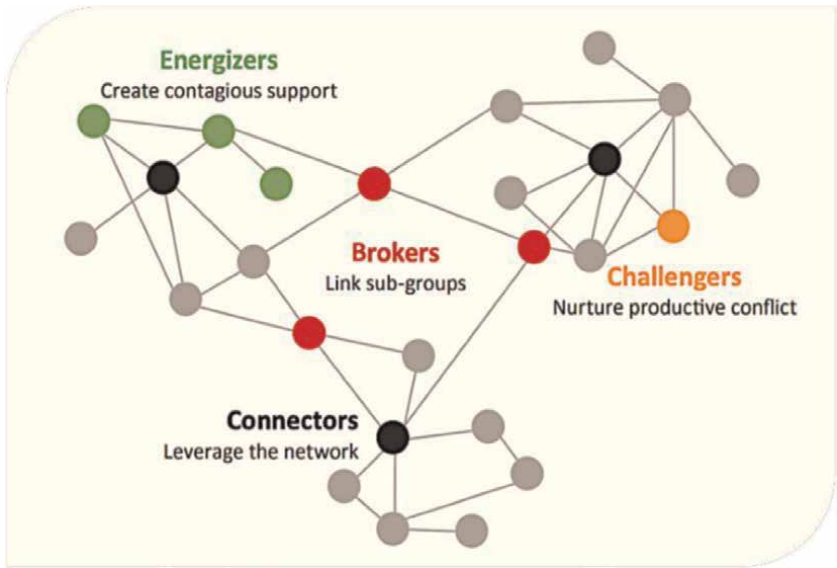
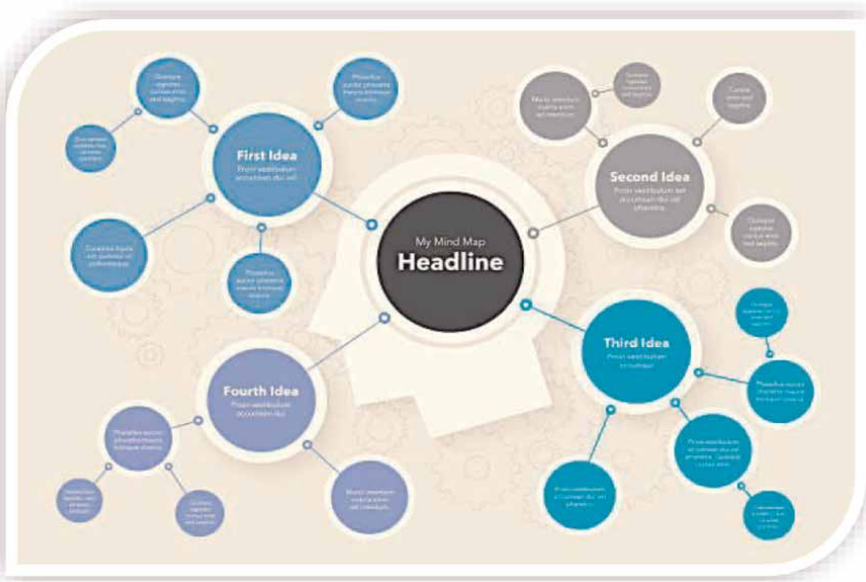


Figure 9.
Social network analysis.



Furthermore, the contribution of the mathematical criterion proposed here to Social Network Analysis covers a vast range of operational applications and could even be employed to determine strategic and tactical actions in crisis areas or to counter terrorist actions (by collecting various sets of information and appropriately analyzing and aggregating them): in general, therefore, this methodology can constitute a valid support for the taking of complex decisions, when it is necessary that the information collected is examined by a human controller.

However, for immaterial networks, unlike those transported by road or rail, to mathematically define the probability of interference between network flows and the “obstacle” along the path which generates the connection defect in the graph, it is necessary to refer to a test-body consisting of a segment and look for a flat geometric figure able to characterize the seriousness of the functional impediment produced.

To this end, a specific degree of impedance of the network was identified (SNA, Mental Maps, etc.), in analogy to the physical quantity that represents the opposition force of an electric circuit to the passage of a variable current, in order to be able to apply the resolution algorithm corresponding to the pre-selected obstacle also to this type of network, according to what is reported in the aforementioned “catalog”, where—case by case—the probability of occurrence of interference is explained.

Based on the above, the catalog already reported for the material networks examined was completed with the subsequent supplementary abacus whereby, for each geometric form of “obstacle” considered therein (triangle, semicircle, square/rhombus, or circle), a precise “degree of impedance” is associated, synthetically defined according to an ordinal scale, formed by three characteristic levels:

1. Elevated (triangle)
2. Middle (semicircle)
3. Low (square or other geometric shape of the “catalog”)

Among intangible networks, one of fundamental importance is the economic-financial network aimed at defining the various logical links between “operators in the sector” for analysis of the type-flows to be studied (securities, etc.). However, the WEB network is also of great importance, considering that, at an international level, even the United Nations Human Rights Council (resolution A/HCR/20/L.13), has expressly considered the Internet a fundamental human right, included in art. 19 of the Universal Declaration of Human and Citizen Rights.

In this context, a performance element of the reticular system is the speed of the network (or of the links) in transmitting and receiving data securely.

In the case of a packet-switched network (finite and distinct sequence of data), the problem generally affects the operation of several networks (**Figure 11**) and the metrics that determine their performance are measured in terms of bandwidth, bit rate, throughput, latency (delays or delays) and packet loss.

In any case, network congestion phenomena must be prevented by monitoring any traffic excesses that could block data paths along the graph representing the network structure.

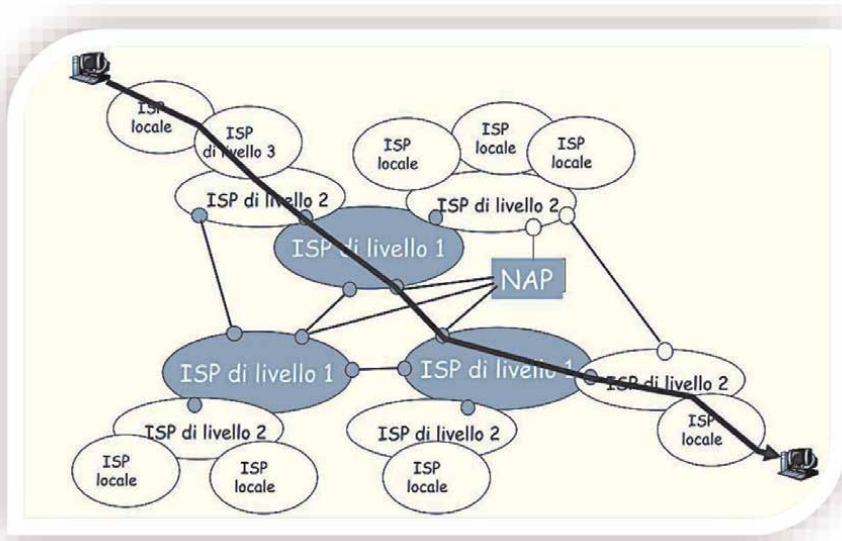


Figure 11.
 Description of the path of a data stream along multiple networks.

4. Road network analysis with GIS and GRASS-GIS

The probability can be useful of intersections in the determination of a classification rule for raster conversions in Geographical Information System (GIS) and GRASS GIS for the Road Network Analysis (RNA) [15, 24–27].

To store the referenced information in some regular lattice covering a projected portion of the Earth's surface. These lattices might be quadratics in an ecological study region, square cells on a raster map, rectangular sheets in a map series, spatial partitions of a large geographical database, or zones in a georeferencing system.

The power of such solutions is to characterize the computational complexity of GIS algorithms and thereby estimate their efficiency [10, 28, 29].

Several papers have employed Buffon's Needle framework to models of migration and travel distances. Rogerson [13] developed intraregional migration distance models; the probability that a move of a given distance will not cross a regional boundary was estimated under several scenarios. Kirby considered trip surveys and the probability of undercounting trips that did not cross survey boundaries.

Geometric probability is a branch of mathematics that is concerned with the probabilities associated with geometric configurations of objects. Among the most famous of these applications is Buffon's Needle problem. The eighteenth-century French naturalist Buffon conceived of the problem while considering a popular game of chance in which a stick was thrown upon a tile floor. The problem continues to appeal to students of mathematics for its variety of elegant solutions [30–32].

In a previous paper Caristi et al. [33] considering the model introduced in the second paragraph, the variations of the classic Buffon's Needle problem are of particular interest. Indeed, if we consider a tile $\mathcal{R}(a, b, \alpha)$ composed of irregular fundamental cells C_0 represented as in **Figure 12**.

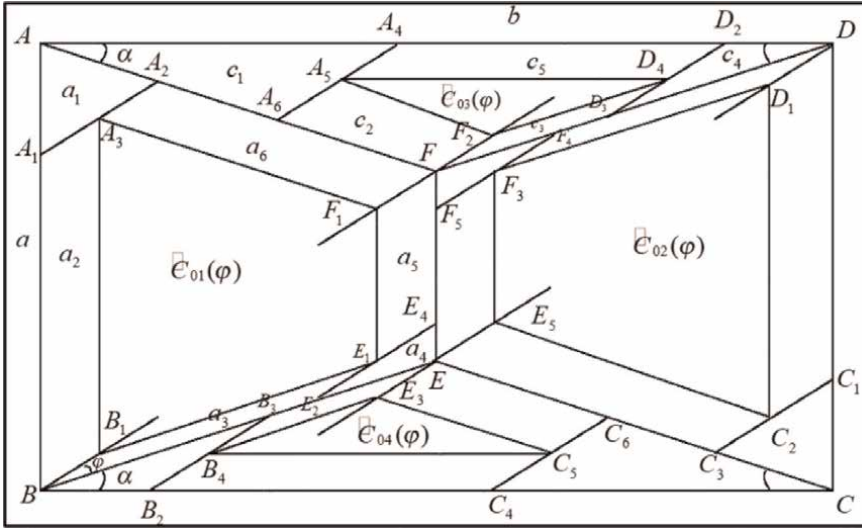


Figure 12.
Lattice of limit positions with segment obstacles.

where $\alpha \in]0, \pi/4]$ is an angle and $a > btg\alpha$.

Denoting with M_1 and with M_2 the set of all segments s that have their centre in C_{01} and C_{03} . Denoting, likewise, with N_1 and N_2 the set of all segment s completely contained within C_{01} and C_{03} . We get:

$$P = \frac{\mu(N_1) + \mu(N_3)}{\mu(M_1) + \mu(M_3)}, \quad (14)$$

where μ denotes the Lebesgue measure in the Euclidean plane [14].

To compute the above measures, we use the Poincaré kinematic measure [26]:

$$dK = dx \wedge dy \wedge d\varphi, \quad (15)$$

where x, y are the coordinates of the point 0 and φ the angle of s .

By **Figure 1**, $0 \leq \varphi \leq \pi/2 - \alpha$, and we have

$$\begin{aligned} \mu(M_1) &= \int_0^{\frac{\pi}{2}-\alpha} d\varphi \iint_{\{(x,y) \in C_{01}\}} dx dy = \int_0^{\frac{\pi}{2}-\alpha} (area C_{01}) d\varphi = \left(\frac{\pi}{2} - \alpha\right) area C_{01}, \\ \mu(M_3) &= \int_0^{\frac{\pi}{2}-\alpha} d\varphi \iint_{\{(x,y) \in C_{03}\}} dx dy = \int_0^{\frac{\pi}{2}-\alpha} (area C_{03}) d\varphi = \left(\frac{\pi}{2} - \alpha\right) area C_{03}, \\ \mu(M_1) + \mu(M_3) &= \left(\frac{\pi}{2} - \alpha\right) \frac{ab}{2}, \end{aligned} \quad (16)$$

and

$$\mu(N_1) = \int_0^{\frac{\pi}{2}-\alpha} d\varphi \iint_{\{(x,y) \in \widehat{C}_{01}\}} dx dy = \int_0^{\frac{\pi}{2}-\alpha} [area \widehat{C}_{01}(\varphi)] d\varphi, \quad (17)$$

$$\mu(N_3) = \int_0^{\frac{\pi}{2}-\alpha} d\varphi \iint_{\{(x,y) \in \widehat{C}_{03}\}} dx dy = \int_0^{\frac{\pi}{2}-\alpha} [area \widehat{C}_{03}(\varphi)] d\varphi. \quad (18)$$

Integrating relations (17) and (18) and substituting in (14) we obtain

$$P = 1 - \frac{4}{(\pi - 2\alpha)ab} \left\{ al(1 - \sin \alpha) - \frac{bl}{2}(1 - \sin \alpha + 3 \cos \alpha) - \frac{l^2}{2} \left[\cos 2\alpha + \left(\frac{\pi}{2} - \alpha \right) (1 + \operatorname{ctg} \alpha) \right] \right\}. \quad (19)$$

Denoting with

$$(\alpha) = \frac{2al(1 - \sin \alpha) - bl(1 - \sin \alpha + 3 \cos \alpha) - l^2 \left[\cos 2\alpha + \left(\frac{\pi}{2} - \alpha \right) (1 + \operatorname{ctg} \alpha) \right]}{\pi - 2\alpha} \quad (20)$$

we can write

$$P = 1 - \frac{2}{ab} f(\alpha). \quad (21)$$

We prove that there exists a system value for α, a, b, l for which the probability P is maximum. Indeed, for $\alpha = \frac{\pi}{4}$ it is easy to verify that $f'(\alpha) = 0$ and $f''(\alpha) > 0$ then the probability P is maximum.

For our considered lattice we have: Example 1. What is the probability that the body test is missed during a 2D conversion to raster?

Theorem 1. The probability P_{int} that a random segment s of fixed length l , fulfilling the relation $l < a/2$, uniformly distributed in a bounded region of the plane, intersects a side of the lattice $\mathcal{R}(\alpha)$ is:

$$P = \frac{4}{(\pi - 2\alpha)ab} \left\{ al(1 - \sin \alpha) - \frac{bl}{2}(1 - \sin \alpha + 3 \cos \alpha) - \frac{l^2}{2} \left[\cos 2\alpha + \left(\frac{\pi}{2} - \alpha \right) (1 + \operatorname{ctg} \alpha) \right] \right\}. \quad (22)$$

Corollary 1. For $\alpha = \frac{\pi}{4}$, by (21) we have that $0 \leq P_{int} \leq \frac{2}{ab} f(\alpha)$.

We consider a tile $\mathcal{R}(a, \alpha, \beta)$ composed of an irregular fundamental cell $C_0 = C_{01} \cup C_{02}$ represented in **Figure 13** where α and β are angles with $\frac{\pi}{4} < \alpha \leq \frac{\pi}{3}$ and $\beta \leq \alpha$ (**Figure 13**):

In the same way, we have: Example 2. What is the probability that the body test rectangle is missed during a 2D conversion to raster?

Theorem 2. The probability P_{int} that a random rectangle r of side a and b , uniformly distributed in a bounded region of the plane, intersects a side of the lattice $\mathcal{R}(a, \alpha, \beta)$ is:

$$P_{int} = \frac{1}{3a^2(\alpha \tan \alpha + \beta \tan \beta)} \left\{ al \left[6 - 4 \cos \alpha - 4 \cos \beta + \frac{1}{\cos \alpha} + \frac{1}{\cos \beta} \right] - \left(2 \sin \alpha + 2 \sin \beta - \frac{l^2}{2} [\sin \alpha (\sin \alpha + \cos \alpha) + \sin \beta (\sin \beta + \cos \beta) - \alpha \cot \alpha - \beta \cot \beta] \right) - \frac{m^2}{2} (\sin^2 \alpha + \sin^2 \beta) - \alpha \cot \alpha - \beta \cot \beta \right\}. \quad (23)$$

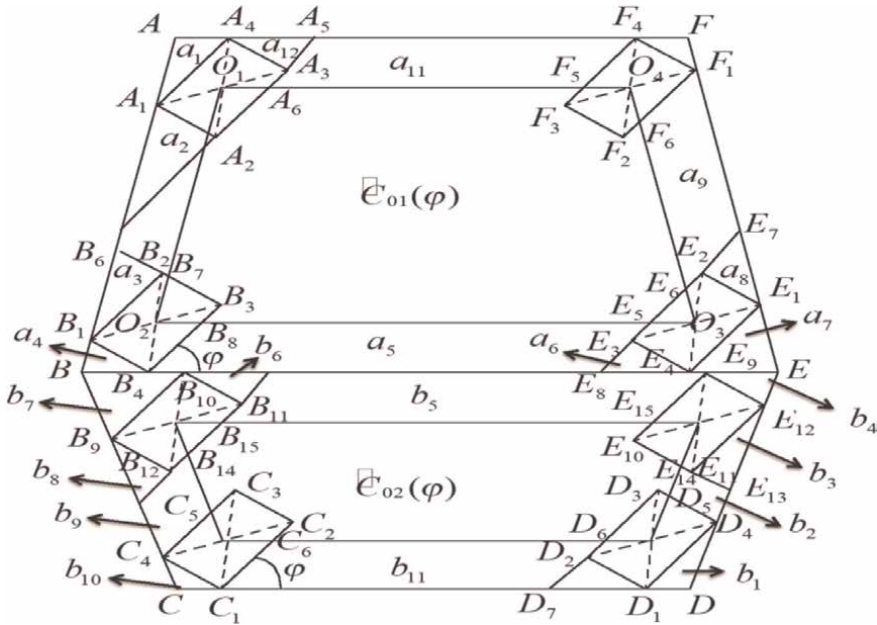


Figure 13.
Lattice of limit positions with rectangle obstacles.

For $m = 0$ the rectangle r became a segment of length l and we find the probability determined in [30].

5. Conclusions

In this chapter we highlight the relevance of geometric probabilities for Road Network Analysis applications. In particular, we show how the probability of intersections and geometric probabilities approach for transportation planning operations can be useful.

To begin with we represent the network with special evaluated graphs, integrated by mathematical models of stochastic geometry and analytically tackled through an appropriate interpretation of the “graph” constructed with “attributes” and thought of as formed by a whole union of regular and/or irregular elementary geometric figures, constituting special lattices R , in the geometric space S^m . Subsequently, we extend the same representation for the Social Network Analysis and in conclusion we show the utility of the probability of intersections in the determination of a classification rule for raster conversions in Geographical Information System (GIS) and GRASS GIS.

Author details

David Barilla¹, Giuseppe Caristi^{1*}, Roberto Guarneri¹ and Dario Lo Bosco²

¹ Department of Economics, University of Messina, Italy

² RFI, Italy

*Address all correspondence to: gcaristi@unime.it

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Belli P, Anderson JR, editors. *Economic Analysis of Investment Operations. Analytical Tools and Practical Applications*. Washington D.C: WBI, World Bank; 2001
- [2] Billio M, Lo AW. *Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors*. Venice: Department of Economics, Ca' Foscari University of Venice, Working Paper Series; 2011
- [3] Dornbusch R, Fischer S, Startz R. *Macroeconomia*. Milano, Italia: McGraw-Hill Companies; 2010
- [4] Fossati A, Pubblica E. Franco Angeli. Italia: Milano; 2002
- [5] Goolchildfd MF. A hierarchical spatial data structure for global geographic information system, CVGIP. *Graphical Models and Image Processing*. 1992;54(1):31-44
- [6] Kohli KN. *Economic Analysis of Investment Projects: A Practical Approach*. Oxford: Oxford University Press for the Asian Development Bank; 1993
- [7] Accademia Nazionale Di Scienze. Lettere ED Arti, Atti, Investimenti e Infrastrutture. Volume della Classe di Scienze, I semestre, parte II. 1999. pp. 73-79
- [8] Layard R, Glaister S. *Cost Benefit Analysis*. Cambridge University Press; 1994
- [9] Jacoby HG, Minten B. On measuring the benefits of lower transport costs. In: *The World Bank*. Washington USA: Policy Research Working Paper 4484. 2008. pp. 1-20
- [10] Higgins PM. *Nets, Puzzles, and Postmen. An Exploration of Mathematical Connections*. Oxford University Press; 2007
- [11] Leonardi G, Moretti M, Stoka M. *L'analisi del funzionamento della rete. Un approccio metodologico*. Rome, Italy: Argomenti, N. 1, R.F.I; 2003
- [12] Solomon H. Geometric probability. In: *CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia: Society for Industrial and Applied Mathematics; 1978. p. 28
- [13] Rogerson PA. Buffon's needle and the estimation of migration distances. *Mathematical Population Studies*. 1990; 2:229-238
- [14] Stoka M. *Probabilità e geometria*. Palermo, Italy: Ed. Herbita; 1982
- [15] Watts DJ. The “new” of networks. *Annual Review of Sociology*. 2004;30: 243-270
- [16] Uspensky JV. *Introduction to Mathematical Probability*. New York: McGraw-Hill; 1937
- [17] Santalo LA. *Integral Geometry and Geometric Probability*. Mass: Addison Wesley; 1976
- [18] Aragone F. *Imparare con le Mappe Mentali*. Milano: Enea; 2015
- [19] Breiger RL. The analysis of social networks. In: Hardy M, Bryman A, editors. *Handbook of Data Analysis*. London: SAGE; 2004. pp. 505-526
- [20] Butt CT. Social network analysis with SNA. *Journal of Statistical Software*. 2008;24(6):1-51
- [21] Carrington PJ, Scott J, Wasserman S. *Models and Methods in Social Network*

Analysis. United Kingdom: Cambridge University Press; 2005

[22] Clarke CK. Analytic and Computer Cartography. 2nd ed. Prentice Hall; 1995

[23] Barabási AL. Link. La Scienza Delle Reti. Torino: Einaudi Editore; 2004

[24] Adhikari A, Using Data Network Metrics. Graphics and topology to explore network characteristics. Networks and Beyond. 2007;54:62-75

[25] Burrough PA, McDonnell RA. Principles of Geographical Information Systems. United Kingdom: Oxford University Press; 1998

[26] Poincaré H. Calcul des Probabilités. 2nd ed. Paris: Gauthier-Villard; 1912

[27] Trobia A, Milia V. Social Network Analysis. Approcci, Tecniche e Nuove Applicazioni. Roma: Carrocci; 2011

[28] Goodchild MF, Shotridge AM. Geometric probability and GIS: Some applications for the statistics of intersections. International Journal of Geographical Information Sciences, London. 2002;16(3):227-243

[29] Mitsova H, Neteler M. Open Sources GIS: A GRASS GIS Approach. 2nd ed. Cambridge, United Kingdom: Kluwer Academic Publishers; 2004

[30] Barilla D, Caristi G, Puglisi A. A Buffon-Laplace type problems for an irregular lattice with maximum probability. Applied Mathematical Sciences. 2014;8(165):8287-8293

[31] Bonanzinga V, Sorrenti L. Geometric probabilities for three-dimensional tessellations and raster classifications. Serie on Advances in Mathematics for Applied Sciences;82:111-122

[32] Boots B. Spatial tessellations. In: Longley PA, Goodchild MF, Maguire DJ, Rhind DW, editors. Geographic Information Systems: Principles, Techniques Issues. 2nd ed. New York: Wiley & Sons; 1998. pp. 503-526

[33] Caristi G, Guarneri R, Lo Bosco S. Road network analysis with GIS and GRASS-GIS: A probabilistic approach. Journal of Geographical Research. 2021;4(4):48-52

Applications of Geotechnologies in the Field of Public Health

Melgris José Becerra and Mariano Araujo Bernardino da Rocha

Abstract

This chapter discusses the role of epidemiology and the importance of spatial analysis in understanding patterns of disease occurrence in human populations. Epidemiologists use inductive and deductive approaches to investigate the relationships between risk factors and health outcomes, using advanced techniques such as factor analysis, multilevel modeling, and causal diagrams. Spatial analysis plays a crucial role in medical geography by identifying a disease's spatial and temporal distribution. Methods such as point, line, and surface patterns are used to analyze spatial clusters, connections, and trends in disease distribution. These techniques provide valuable information for public health decision-making. The COVID-19 pandemic has highlighted the importance of spatial analysis, using geographic information systems and web-based tools to track the spread of the virus. Advances in geoprocessing techniques, particularly geographic information system (GIS), have transformed medical geography. GIS makes it possible to describe, analyze, and predict spatial patterns by integrating data from different sources and spatial scales. These tools facilitate the creation of spatially explicit variables and allow the exploration of spatial groupings and relationships in the data. Considering the spatial context is essential to understand the determinants of health and the influence of the environment on individual and collective health.

Keywords: epidemiology, spatial analysis, medical geography, geographic information system, geoprocessing

1. Introduction

Epidemiology is commonly regarded as the methodological backbone of the fields of public health and outcomes research, as its practitioners describe patterns of disease occurrence, identify risk factors and etiological determinants, and demonstrate the utility of interventions [1]. Epidemiological research has made numerous discoveries that have contributed to improving public health. For example, epidemiologists have identified many health risks from active and environmental exposure to tobacco and have found effective smoking prevention and cessation options.

Epidemiologists are constantly seeking to improve their understanding of patterns of disease occurrence in human populations [2]. Historically, they have used inductive approaches to identify simple relationships between risk factors and health outcomes.

However, many of the more complex relationships between risk factors and outcomes have yet to be discovered. To address this challenge, techniques such as factor analysis, multilevel modeling, and causal diagrams have been developed to help understand these complex relationships. Although these techniques have limitations in their ability to uncover previously unimaginable interrelationships between etiologic factors, they have been valuable in advancing epidemiologic understanding [3]. In addition, epidemiologists have also adopted deductive approaches, such as dynamical systems models, which allow the prediction of empirical observations and can be particularly useful in the study of infectious diseases [4]. These advanced mathematical models can accelerate hypothesis generation and improve our description of a complex world, allowing epidemiologists to investigate interactions more efficiently between system elements and establish new hypotheses for further testing and validation.

1.1 Importance of medical geography in the spatial analysis of disease

Spatial analysis techniques, such as point, line, and surface patterns, play a fundamental role in medical geography for disease analysis [5]. These techniques provide tools to identify the spatial and temporal distribution of diseases, which allows an understanding of the patterns of occurrence. Point pattern methods allow for examining the spatial clustering of disease occurrences, while line methods analyze spatial connections and relationships [6]. On the other hand, surface techniques are used to identify trends and gradients in disease distribution. In addition, comparison methods and autocorrelation measures are available to help assess spatial similarity between different geographic areas. These spatial analysis techniques are fundamental to understanding the geography of disease and provide valuable information for public health decision-making.

The COVID-19 pandemic has further highlighted the importance of spatial analysis in the health field. Maps have been widely used to track the spread of the virus and understand its spatiotemporal diffusion. The use of geographic information system (GIS) and web-based tools has revolutionized health geography by providing the ability to analyze and visualize disease patterns more efficiently. However, spatial analysis of epidemiological data in real-time presents significant challenges. The speed with which data are generated for an emerging disease such as COVID-19 requires agile technical and methodological solutions. Despite advances, challenges exist in real-time data integration, interpretation of results, and implementation of measures based on spatial analysis. As we face these challenges, it is critical to take full advantage of spatial analysis techniques to effectively understand and address diseases in the geographic context.

1.2 Evolution of geoprocessing techniques and their influence on the transformation of medical geography

Advances in geoprocessing techniques, especially geographic information system (GIS), have had a significant impact on the transformation of medical geography [7]. GIS has been widely used in epidemiological applications, enabling the description, analysis, and prediction of spatial patterns using geographic data and related attributes. These tools have facilitated the creation of spatially explicit variables, such as measures of the built environment, environmental exposures, and demographic characteristics [8]. In addition, GIS has enabled the integration of data from different sources and spatial scales, which has improved the visualization, management, and analysis of information.

GIS-generated thematic maps have been valuable for identifying high-risk areas, highlighting health disparities, assessing resources, and generating explanatory hypotheses [9, 10]. From an analytical perspective, GIS tools have been used to explore spatial or spatiotemporal groupings, investigate locally varying relationships, and model spatial dependencies in data [11]. These analytical resources have provided a deeper understanding of spatial factors and their influence on population health.

It is critical to consider the spatial context in which people live and seek care. The location and characteristics of the environment can influence health outcomes at the individual level, as well as access and quality of care at the provider and system level. Contextual data are important for both the description and analysis of health determinants. For example, one can examine the distribution of green space in a region and develop an index that predicts the likelihood of meeting physical activity guidelines or individual weight status. In addition, hierarchical models allow one to examine how outcomes vary at the individual and group levels, as well as to identify external factors that affect people's health and well-being. It is essential to consider the role of "place" in epidemiological analysis, as it can have a significant impact on health and should be considered in health interventions and policies.

2. Health data processing and geotechnologies

2.1 Definition of geoprocessing and its main applications

"Geoprocessing is a set of computational techniques that operates on georeferenced databases, which are records of occurrences, to transform them into relevant information" [12]. But what is this relevant information? This question is important to understand a type of information called geoinformation and, consequently, other terms related to geoprocessing, such as geoinformation science, geographic information system (GIS), remote sensing (RS), digital image processing (DIP), geostatistics and spatial analysis.

Currently, we live, according to Takahashi [13], in a unique moment for the world society, where information flows at unimaginable speed and quantity, and the Internet is the icon of this society, known as the Information Society.

Information is generated by the processing or interpretation of data, which is the symbolic form of representation of reality in a quantified way. For example, when we write the letters "A" "D" "A" "T", which are symbols of grammatical reality in the English language, they have no meaning. However, when we rearrange these same letters to form the word "D" "A" "T" "A", we obtain a word with four letters and a meaning. In this moment of processing the word DATA, information is generated.

In the universe of geoprocessing, we deal with a special type of data, and we generate very particular information, namely the geographic data and the geoinformation. Therefore, it is always necessary to work with a location in space. Geographic data are those that intrinsically have location, such as latitude and longitude, and geoinformation or spatial information, which is the result of processing spatial data manipulated by computers, maintaining the location in space.

Geoprocessing and remote sensing are two disciplines or technologies that provide the basis for geoinformation science and are equipped with computational tools, such as GIS and PDI, respectively. These technologies, together with geostatistics and spatial analysis techniques, form a set called geotechnologies. **Figure 1** presents a visualization of the hierarchy and connection between these concepts, and **Table 1** presents the concepts of each of these terms based on prominent sources.

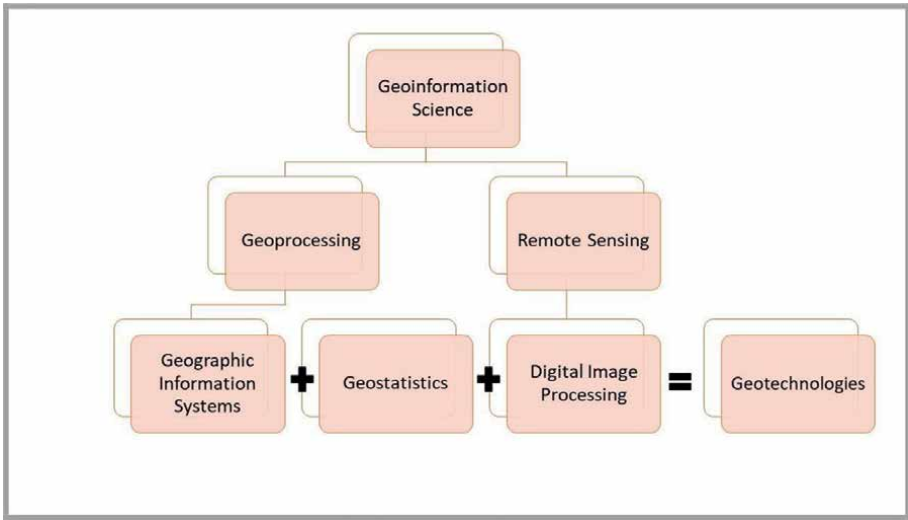


Figure 1.
Relationship between various terms related to geoprocessing. Source: The authors.

Terms	Concepts	Sources
Geoinformation science	Study and implementation of different forms of computational representation of geographic space.	Câmara et al. [14]
Geoprocessing	Set of technologies that enables the manipulation, analysis, simulation, and visualization of georeferenced data.	Fitz [15]
Remote sensing	Technology of obtaining data from the Earth's surface as images by capturing and recording electromagnetic energy emitted or reflected by the Earth's surface.	Jensen [16]
GIS	Set of hardware, software, and peopleware for capturing, analyzing, storing, managing, and presenting geographic data.	Silva [17]
PDI	It is a sequence of computational processes that allow the rapid manipulation of matrix data for maximum information extraction. It is viewed in terms of one basic requirement, which is interactive (near real-time) operation, and two general characteristics, which are high data volume and the repeated application of a sequence of calculations to each data element.	Fuat Ince [18]
Geostatistics	Subfield of statistics that studies regionalized variables.	Yakamoto and Landim [19]
Spatial analysis	Set of procedures for the purpose of choosing an inferential model that explicitly considers the spatial relationship present in the phenomenon.	Druck et al. [20]
Geotechnologies	These are sets of scientific techniques and methods applied to the analysis, exploration, study, and conservation of natural resources, considering different scales and spatial information (geographic location).	Tôsto et al. [21]

Source: The authors.

Table 1.
Definitions of terms related to geoprocessing.

2.2 Introduction to geographic information systems (GISs) and their functionalities

The geographic information system (GIS) emerged in the 1980s, when the price of computers began to decline enough to sustain the software industry, considering the cost-benefit of its applications [22]. The timber sector was the first to experiment with this technology, due to the need to control its resources and regulate their use. After this moment, the evolution in the technological, academic, and commercial fields allowed the GIS to be improved according to the new conditions imposed. The launch of satellites, the availability of satellite positioning systems, the internet, and more robust software and hardware have contributed to this improvement.

GISs are powerful tools and services for socio-environmental planning. This system aims to process a large amount of data that describes the environment in an advanced way, both in ordinary and extraordinary situations. The combined use of this tool allows us to generate useful scenarios for decision-making regarding land management [23].

GISs are systems that combine the semantic component of land features (mainly tabular data) with their georeferenced geometric representation. They are equipped with various processors and tools that allow the effective representation of the territory. For example, digital surface models (DSMs) allow the creation of 3D virtual worlds with a major impact on communication.

Like any system, GIS is composed of interrelated elements. In this case, the elements are the software, which provides the functions and tools necessary to store, analyze, and visualize geographic information; the hardware, which consists of the set of equipment necessary for the software to perform its functions; and the peopleware, which refers to the people who, in addition to knowing geoprocessing techniques, must know the application area to correctly select the variables that will simplify the analyzed reality.

According to Magalhães et al. [24], GISs are composed of procedures for obtaining and manipulating graphic and non-graphic data, responsible for data entry, database management systems, spatial data analysis techniques, procedures for data storage and retrieval, and procedures for making results available (visualization, plotting and reporting). What characterizes this system and makes it unique is the possibility to combine geographic data with alphanumeric data, as can be seen in **Figure 2**.

In addition to the function of data storage and information dissemination, GIS also requires the data acquisition stage. This phase is comprehensive, as it involves several techniques. Remote sensing is one of them, which aims to acquire information about targets on the Earth's surface without direct contact with the object. This acquisition can occur passively or actively, and the resulting products are satellite or radar images, as well as graphs, spectral curves, and others. Another form of data acquisition is through satellite positioning systems, whose function is to collect a pair of coordinates to locate specific events (**Figure 3**).

Ramvalho et al. [25] emphasize that geographic information system should not be confused with the same concept of geoprocessing. The latter covers various technologies for the treatment and manipulation of geographic data through computer programs. GISs (geographic information systems) are only tools for the realization of geoprocessing, as illustrated in **Figure 4**.

Therefore, this system, with its multiple facets, both in the production of maps and in solving geographic problems and supporting spatial decisions, presents



Figure 2.
The image in the background shows the location of dams in the Amazon through the blue vertical bars, and the image in the center is a zoom that allows the visualization of the link between alphanumeric data (tabular) and a spatial data (point geometry) highlighting a mining tailings dam in the Amazon. Source: Adapted from <https://tailing.grida.no/map/data/and> and <https://globaia.org/water-anthropocene>.

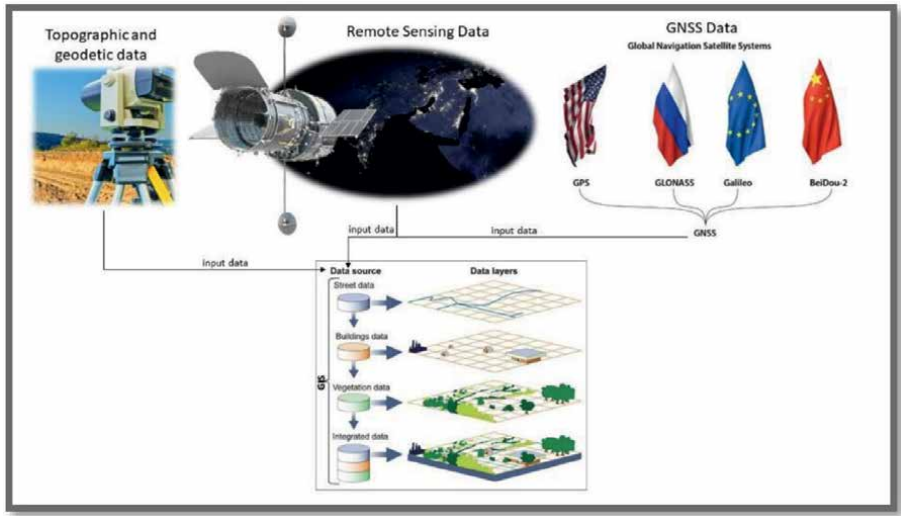


Figure 3.
Example of data input into GIS. Source: Adapted from OXTS (2020) and National Geographic (2023).

elements of paramount importance when connected. A crucial feature is the interoperability of data, allowing the integration of information from different sources and a rapid dissemination of information. The Internet is the available means of diffusion, being responsible for the dissemination and retrieval of data, with numerous websites specialized in geotechnologies [22].

This may seem abstract, but GISs are already part of a technological frontier that is their use on the Internet, through the so-called WebGIS. Among the various examples that could be cited, the most famous are Google Maps and Google Earth, which are

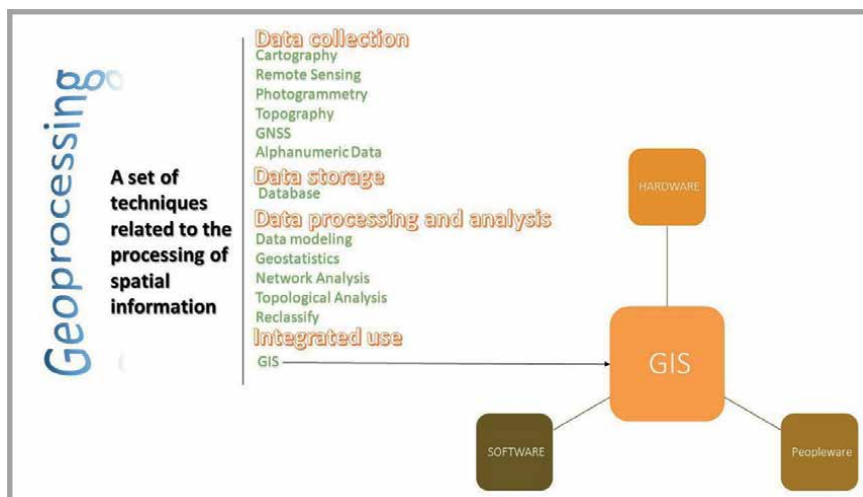


Figure 4.
 Difference between geoprocessing and GIS. Source: The authors.

certainly used in the daily lives of the world's population when they are moving around, either in cities or on trips.

2.3 Using remote sensing to obtain geospatial data

The use of remote sensing to obtain geospatial data is essential in the field of geoprocessing. Geoprocessing and remote sensing are disciplines/techniques that are increasingly interconnected, serving as tools for interdisciplinary research. This research has intrinsic characteristics of dialog and integration between several scientific disciplines [26].

The products generated by the remote sensing tool, through digital image processing, are used as input in geoprocessing through the geographic information system (GIS). These products are mostly matrix data known as satellite images. These images are obtained by sensors attached to orbital satellites, which have radiometers capable of collecting electromagnetic radiation that interacts with targets on the Earth's surface. From these data, it is possible to infer values of variables, such as temperature and air humidity, distinguish different targets on the Earth's surface, and obtain digital terrain models [27].

Remote sensors can be imagers or not and can only generate graphics. They can be manual, such as handheld spectroradiometers, or be attached to balloons, drones, aircraft, and satellites. These sensors can be passive or active (**Figure 5**), collecting part of the wavelengths of electromagnetic radiation emitted by the Sun and interacting with the Earth's surface, or emitting their own pulses in the microwave range and interacting with terrestrial targets, returning to the sensor (**Figure 6**).

To infer values of environmental variables from the Earth's surface, orbital remote sensors use characteristics of electromagnetic waves. They capture fractions of the length of these waves through bands (channels) in the sensors. These fractions are recorded as electromagnetic radiation flux, also known as radiance, and are stored as digital numbers (DN) in a matrix. This matrix is further processed to form an image with understandable data of an environmental variable, which will serve as input for geoprocessing (**Figures 7 and 8**).

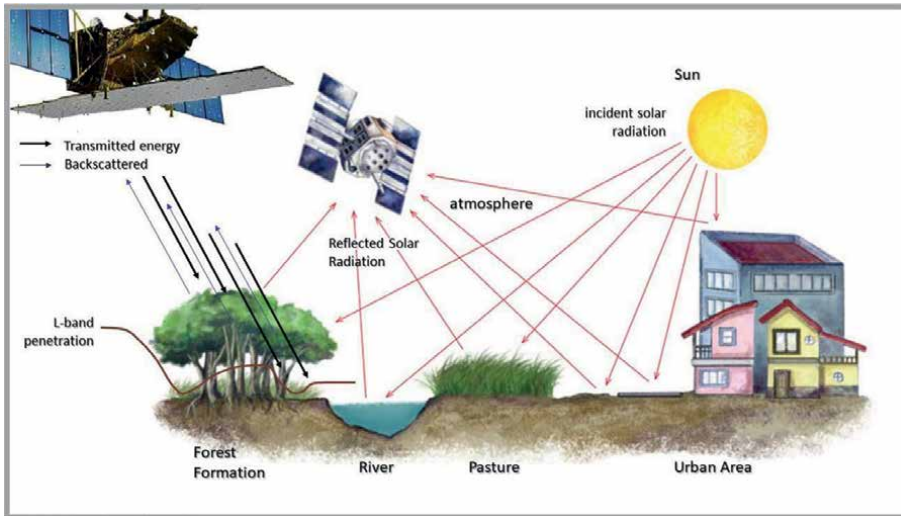


Figure 5. Active sensors (L-band of a synthetic aperture radar - SAR) emitting and receiving pulses and passive sensors, only receiving the radiation from the sun that interacts with the targets. Source: Adapted from IBGE [28].

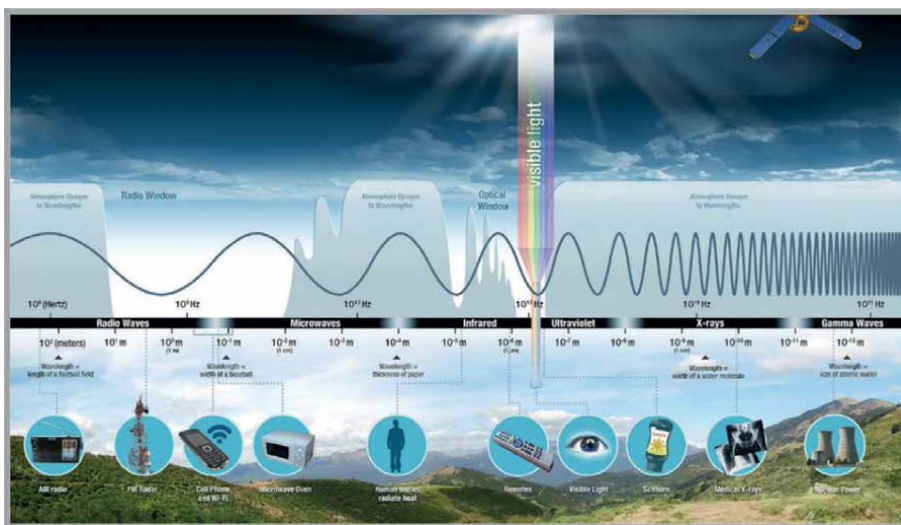


Figure 6. Electromagnetic spectrum with its divisions by wavelength and frequency, the wavelength increases in the direction from right to left of the viewer, and the frequency increases in the opposite direction. Source: <https://science.nasa.gov>.

Thus, orbital remote sensors use electromagnetic radiation to obtain information about the Earth's surface. By capturing this information through different spectral bands, they record the radiation as digital values that can be processed to form images that represent environmental variables. These images are fundamental for geoprocessing, providing relevant data for spatial analysis and decision-making in several areas, such as environmental monitoring, and urban and agricultural planning, among others.

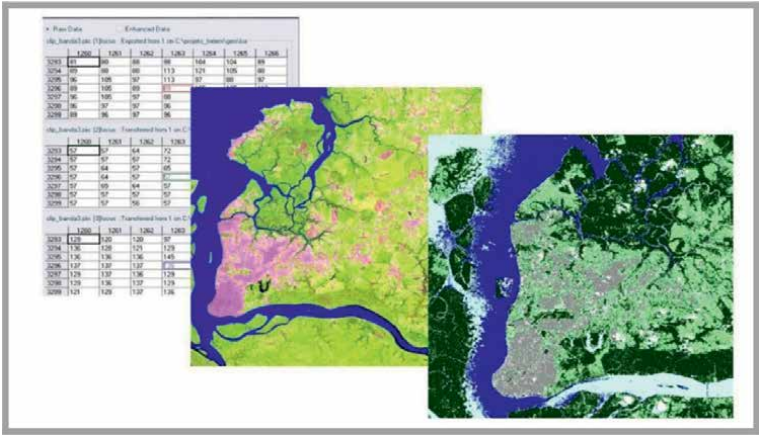


Figure 7.
Raw data from passive sensors with DN in a matrix, forming an RGB color composite image and being processed, generating spatial land use and land cover information. Source: The authors.

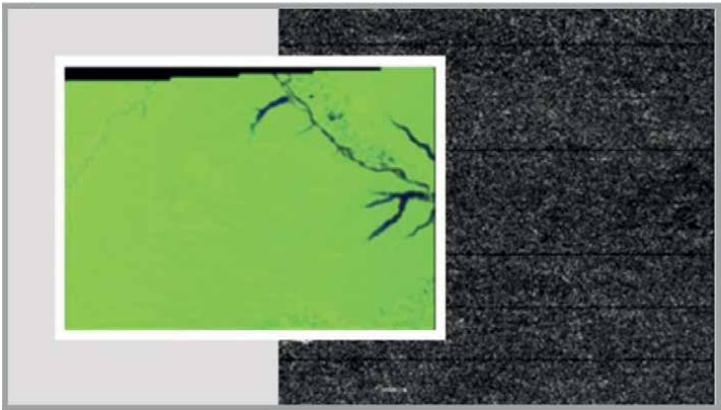


Figure 8.
Raw radar data and its visualization after processing. Source: Maciel and Diniz [29].

Figure 9 shows passive sensors, such as the advanced baseline imager (ABI) of the GOES-16 satellite, which aims to infer terrestrial meteorological data. Bands 1, 2, and 3 of this sensor store data in bands of the visible and near-infrared electromagnetic spectrum. This curve is shown in red, where the peaks represent high transmittance, i.e., a good passage of the wavelengths of radiation that interacts with the target and is recorded by the sensor. The valleys in the curve show high absorption due to the presence of atmospheric gases such as ozone and water vapor.

These passive sensors, such as the ABI, are fundamental for the collection of terrestrial meteorological data. They allow the observation and recording of different wavelengths of electromagnetic radiation, providing information on atmospheric conditions and allowing more accurate weather analysis and forecasting. The use of these passive sensors is essential in monitoring and studying weather and climate, contributing to the understanding of meteorological phenomena and their impacts on the environment and society.

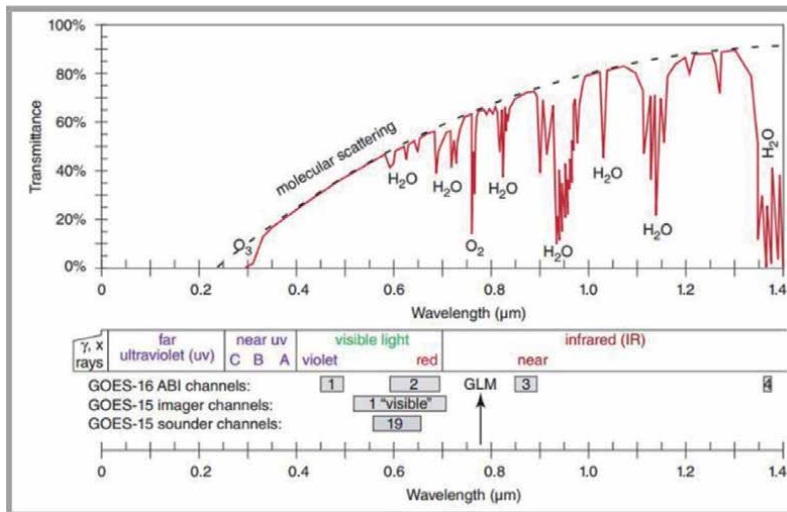


Figure 9.
Atmospheric transmittance of upward electromagnetic radiation from the Earth's surface. Source: [27].

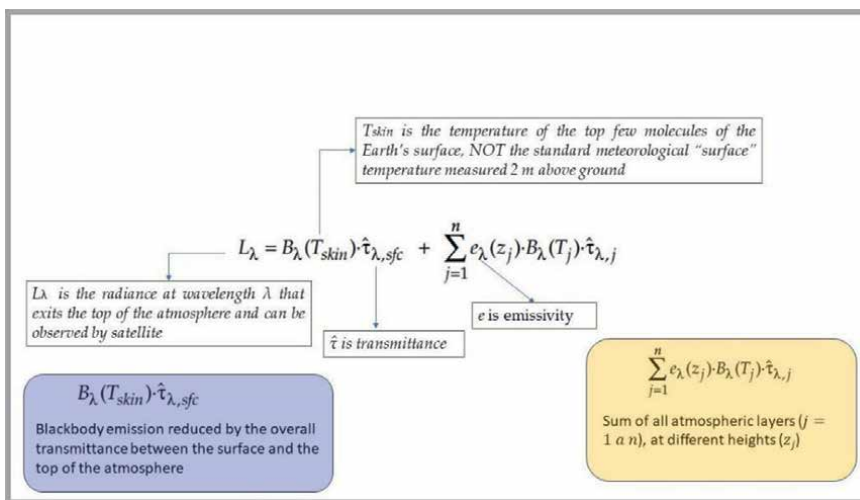


Figure 10.
Explanation of obtaining the radiation flux at one wavelength with an example of the atmospheric interaction based on the Schwarzschild equation. Source: Adapted from [27].

The Schwarzschild equation shows how radiance, one of the sensor inputs, is obtained. In this case, it is used as an example to illustrate the interaction with the atmosphere, as can be seen in **Figure 10**.

The Schwarzschild equation is a fundamental tool in the study of electromagnetic radiation and its interaction with the Earth's atmosphere. This equation allows the calculation of radiance, which is the amount of radiation emitted or reflected by an object in a specific direction.

Figure 10 graphically represents this interaction between radiation and the atmosphere. It can be seen how solar radiation strikes the atmosphere and some of it is

absorbed, scattered, or reflected before reaching the earth's surface. This interaction with the atmosphere has an impact on the data that are captured by the sensors since the radiation undergoes modifications on its way through the atmosphere.

Understanding and modeling this interaction is essential in the processing of remotely sensed data, as it allows correcting for atmospheric effects and obtaining more accurate information about the Earth's surface.

2.4 Collecting and organizing health data

The merging of geographic data with health goes back a long way. This combination of information has been recorded since the nineteenth century, when the English physician John Snow mapped cholera cases and water supply locations in the city of London, England. Although he used methods that are considered simplistic by today's standards and did not rely on the use of hardware and software due to the lack of these technologies at the time, it was possible to overlay the data and visualize, through spatial analysis, the relationship between contaminated water wells and disease outbreaks that resulted in many deaths in the region. This allowed the implementation of public health measures, such as the closure of these wells, to contain the epidemic.

Nowadays, with the advancement of technology, the process of building maps and making decisions has become much faster and more efficient. The possibility of integrating spatial data with quantitative information on diseases and health care has been of great importance. A notable example was the real-time and online updating of data related to the COVID-19 pandemic. As the disease spread around the world and countermeasures were implemented, this data was constantly updated by various professional organizations around the world. **Figure 11** illustrates an example of this type of data visualization.

The integration of geospatial data and health information has led to significant advances in disease monitoring and understanding, enabling better decision-making and the development of effective prevention and control strategies. This collaborative approach between the fields of geotechnology and health has proven increasingly important in the search for solutions to public health challenges.



Figure 11. COVID-19 dashboard: Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Available via the link: <https://www.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6>.

2.5 Available health data sources

The main sources of health data are obtained through health information systems (HIS), which are digital systems that collect data from different sources and make them available through information technology tools to generate strategic information for the benefit of public health. These systems enable decision-makers at all levels of the health system to identify problems and needs, make evidence-based decisions about health policies, and optimally allocate limited resources [30]. Some examples of these systems around the world can be seen in **Table 2**.

Health information systems collect data from different sources, such as medical records, health surveys, epidemiological reports, and mortality records. These data are collected and processed in a standardized way to ensure quality and interoperability. In addition, anonymization and data protection techniques are used to ensure the confidentiality and privacy of individuals' health information.

The availability of these health data sources has enabled significant advances in research, epidemiological surveillance, and health decision-making. The data collected are used to identify disease patterns, evaluate the effectiveness of public health interventions, monitor the spread of infectious diseases, and assess the burden of disease in different populations.

Health system	Characteristics	Country
National Health Service (NHS)	This system is made up of other systems such as Electronic Health Records (EHR), Summary Care Records, and NHS Spine, which is a secure network for sharing patient information between different healthcare providers.	United Kingdom
Veterans' Health Information Systems and Technology Architecture (VistA)	Electronic health record system developed by the U.S. Department of Veterans Affairs (VA). It is used to manage VA health care delivery and patient information.	USA
Sistema Único de Saúde (SUS)	The national public health system has DATASUS (Department of Informatic SUS), a health information system. DATASUS collects and manages health data from across the country, supporting health surveillance, epidemiology, resource management, and monitoring of health indicators.	Brazil
My Health Record	It is a digital health record system implemented to enable individuals to securely store and manage their health information, which can be accessed by health service providers to support coordinated care and improve health outcomes.	Australia
National Health Information System (NHIS)	Provides a platform to capture, store, and analyze health data at the national level. NHIS supports health planning, monitoring of health services, and resource allocation.	Ghana
National Electronic Health Record System (NEHRS)	It is a comprehensive health information system implemented to facilitate the sharing of electronic health records between different health service providers, supporting care coordination and enabling health professionals to access patient information whenever needed.	Singapore

Source: The authors.

Table 2.
Different health information systems implemented around the world.

Importantly, access to and availability of health data varies between countries and jurisdictions, and it is critical to establish policies and legal frameworks that promote the ethical and safe use of these data. Collaboration between health professionals, researchers, and decision-makers is essential to make the most of these data sources and use them to improve the health of populations.

These various health information systems, when combined with GIS, demonstrate their full potential in generating spatial information in the field of public health, enabling more informed decision-making.

2.6 Construction of GIS in health

Pina [31] outlines the steps for building a GIS developed for the healthcare sector. However, the author emphasizes that creating a GIS is not a standalone task for the healthcare sector, as obtaining spatial data requires contacting municipal or state planning agencies. These data include information related to infrastructure, basic sanitation, road networks, and aerial imagery, among others. As data to be analyzed, one can utilize information from available information systems stored in databases, such as the DATASUS in Brazil, which is a database of the Unified Health System (SUS) that provides products and services specifically aimed at the public health sector.

In terms of GIS software for data analysis, there are several options available, both open-source and proprietary. Each has its own advantages and disadvantages, which should be assessed during the system's design phase for the healthcare field. Among the open-source software licensed under the GNU General Public License, one can mention QGIS, driven by volunteers from around the world. On the other hand, among proprietary software options, ArcGIS Pro by Esri stands out as a global giant in this market niche, with its slogan "The Science of Where." Esri developed the shapefile format in the early 1970s, which revolutionized the storage and processing of geospatial data.

However, these two elements alone are not sufficient for the construction of a GIS, as mentioned earlier. Acquiring computers with sufficient processing power to handle the volume of data entering the system is of great importance. Currently, with the vast amount of available data, one frontier of this technology is the use of cloud-based supercomputers, such as Google's Application Programming Interface (API), the Google Earth Engine. This enables working with petabytes of geospatial data [32]. Additionally, skilled individuals in the fields of geoprocessing, remote sensing, and programming are essential to ensure the successful construction of a GIS project.

3. Software for geospatial analysis in health

3.1 Geographic information system (GIS) software. Presentation of the main GIS software used in health

There are many GIS software options available, and each one has tools that can be used in health applications. However, some of the most used software in this field can be found in **Table 3**.

GeoDa is a free and open-source software designed to facilitate spatial analysis and data modeling.

Program	WebLink
GeoDa	http://geodacenter.github.io/index.html
SaTScan	https://www.satscan.org
Epi Info 7	https://www.cdc.gov/epiinfo
MapWindow GIS	https://www.mapwindow.org
QGIS	https://qgis.org
Arcgis Pro	https://www.esri.com
<i>Source: The authors.</i>	

Table 3.
Examples of GIS software.

SaTScan™ is a free software specifically focused on spatial statistics in the health domain. While it can be used in other areas, it is particularly well-suited for health-related studies and is supported by international foundations and institutes in the health field.

Epi Info 7™, developed and maintained by the CDC (Centers for Disease Control and Prevention) in the United States, is a public software specifically designed for health applications. It offers mobile, desktop, and cloud versions, allowing the generation of input databases through forms and enabling spatial analysis, mapping, and planning of epidemiological surveillance systems.

MapWindow GIS is an open-source desktop GIS software project that provides flexibility for programming and the deployment of plugins.

QGIS is the most widely used free and open-source GIS software worldwide. Maintained by volunteers and supported by donations, it offers numerous plugins and programming capabilities, including Python scripting, attracting a large user community.

ArcGIS Pro is a proprietary software that allows customization through open-source code, enabling programmers to develop specific plugins for their work. It is a leading software in the industry, although its high cost should be considered during the planning of the desired solution.

Additionally, **Table 4** presents some global geospatial databases in the health field.

3.2 Spatial analysis tools

The description of the main tools and algorithms used in spatial analysis aims to identify existing patterns in geographic data and determine relationships between different spatial variables. For this purpose, the tools used in geoprocessing for spatial analysis are grouped into Spatial Statistics, Geostatistics, Uncertainty Representation, and Dynamic Modeling [20].

Other classifications of spatial analysis tools can be found in the literature, based on data models such as points, lines, areas, tabular data, networks, and images [33]. However, Rosa [34] highlights the most common spatial analysis operations in GIS, which can be subdivided into reclassification, overlay, weighting, cross-tabulation, neighborhood analysis, network analysis, and querying georeferenced databases.

Reclassification is performed to improve the visual representation of data or when the attributes of the original data are not suitable for the analysis of the specific problem.

Database	WebLink
The Global Health Observatory	https://www.who.int/data/gho/data/indicators
Global Health Data Exchange	https://ghdx.healthdata.org/
Organization for Economic Co-operation and Development	https://stats.oecd.org/Index.aspx?DatasetCode=HEALTH_STAT
Healthsites	https://www.healthsites.io/map
Eurostat	https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/healthcare-services

Source: The authors.

Table 4.
Geodatabase.

For example, in the MapBiomass project (<https://mapbiomas.org/>), the categories of land use and land cover in the Amazon region can be too detailed for an epidemiological study that requires relating a specific vector with the environment. In such cases, the land use and land cover categories can be grouped into a smaller number, indicating areas with low, medium, and high risk of contact with the vector.

Overlay is often used when extracting data from two or more maps, such as determining areas where a specific land use type occurs on certain soil types. Soil information is contained in one map, while land use information is in another. This function allows for analysis based on a Boolean approach.

Weighting involves generating a weighted average map based on multiple information layers in the database, such as soil, topography, geology, and vegetation cover. A typical example of weighting is the creation of agricultural suitability maps using various information layers.

Cross-tabulation calculates the intersection areas between classes from two information layers, determining the areas of intersection. **Table 5** presents the results of a cross-tabulation between slope and land use maps in a specific watershed.

These spatial analysis tools play a crucial role in uncovering insights and relationships within geographic data, aiding decision-making processes in various fields, including public health and epidemiology.

Neighborhood operators select an area located at a certain distance from a feature of interest. Generally, the user can specify a distance criterion (radius, elevation threshold, etc.) from an object within the information layer, and this function generates a new information layer containing the impact zone around the selected object. For example, from the information layer containing contour lines of a specific area, a new information layer can be generated using this function to identify areas within a certain altitude range. An example of applying the neighborhood operation is the determination of environmental protection areas around watercourses and reservoirs.

Network analysis encompasses a variety of types, and its main applications include determining the shortest path, flow functions, and hydrology. A complex network analysis function involves delineating drainage basins by calculating the direction of surface flow using topographic maps. Another example is evaluating alternative routes for emergency vehicles based on a combination of total path distance and anticipated congestion, which is dependent on the time of day.

Querying a georeferenced database, using the key attribute of GIS, which is the linkage of spatial variables with alphanumeric data stored in a geographic database,

Land Use and Land Cover Category	(%)			
	<2	2 a 5	5 a 10	>10
Forest	549	96,75	9,25	—
Savannah	392,5	2,75	1,75	20,05
Pasture	29,969,75	9207,75	3537,25	565,75
Reforestation	9804,75	1247	188,5	13,5
Perennial Crop	254,25	55,25	—	—

Source: Adapted from [34].

Table 5.
Result of the cross-tabulation between two information layers (in hectares).

allows for queries using SQL (Structured Query Language) in many software applications. For example, using a map and a georeferenced database organized by municipalities, one can answer questions such as: Which municipalities have a population density equal to or greater than 5 inhabitants per square kilometer? Locate on the map the cities located within 60 km of municipality X. What is the distance between cities X and Y?

Geographic analysis techniques consist of primitive spatial operations that can be applied to one or more information layers with the aim of creating new information layers or calculating measures. The combination of these operations is mostly performed using Boolean algebra.

3.3 Exploration of clustering techniques, density analysis, and spatial interpolation

Some spatial statistics techniques in geoprocessing allow for confirmatory analysis, verifying the existence of perceived clusters, or exploratory analysis, searching for cluster patterns. Clustering methods can provide answers to a range of public health questions, such as: Are there unusual clusters of health events in an area? Which locations have a high prevalence of a certain disease? Where are the risks of a disease higher or lower?

Cluster detection methods can be divided into three groups: global methods, which assess clustering in the entire study area; local methods, which identify the location of clusters; and specific methods, which evaluate clustering around a point source, such as a hazardous facility.

The local measure of spatial autocorrelation, known as LISA (Local Indicators of Spatial Autocorrelation), is a measure of clustering that relates a particular location to the values of adjacent areas. Conversely, the Moran's Index (a global method) can be used to assess whether the sampled data exhibit spatial autocorrelation across the entire area. The LISA statistic for subareas i is calculated using Eq. (1).

$$I_i = Z_i \sum_{i \neq j} W_{ij} Z_j \quad (1)$$

Where z_i represents the normalized value of the health indicator of interest in zone i , and w_{ij} is a spatial weight that measures the proximity between zones i and j [35].

Figures 12 and 13 provide an example of the Moran's Index and LISA, respectively, conducted by Anselin [36], using data from a classic social science study by Andre-Michel Guerry on crime, suicide, literacy, and other “moral statistics” from 1830 in France. In this specific case, a univariate analysis was performed with the per capita charity donation data (Donatns variable) yielding a Moran's Index of 0.353, which ranges from -1 to 1 . The closer it is to zero, the lower the spatial auto-correlation. The LISA presents a default result for 999 permutations with a p-value of 0.05, with the significance mapping shown on the left and the cluster mapping on the right.

The Kernel density estimation (heat maps) is not a clustering process itself, but a method for exploring and visualizing patterns of spatial point data presented in concentration areas. As explained by Rizzati et al. [37], it involves transforming point data into a continuous analysis surface through interpolation, allowing for the examination of potential data trends even in areas where no actual data has been collected.

Figure 14 visually demonstrates the principle of Kernel density estimation, and Eq. (2) serves as the basis for calculating the smoothing of probabilities.

$$\lambda_r(P) = \sum_{i=1}^n \frac{1}{R^2} k\left(\frac{P - P_i}{R}\right) \quad (2)$$

Where λ is an estimator that depends on P (locations), R is the radius of influence, “ n ” is the events, and “ k ” is a bivariate kernel function (which can be quadratic, triangular, uniform, Epanechnikov, or Gaussian).

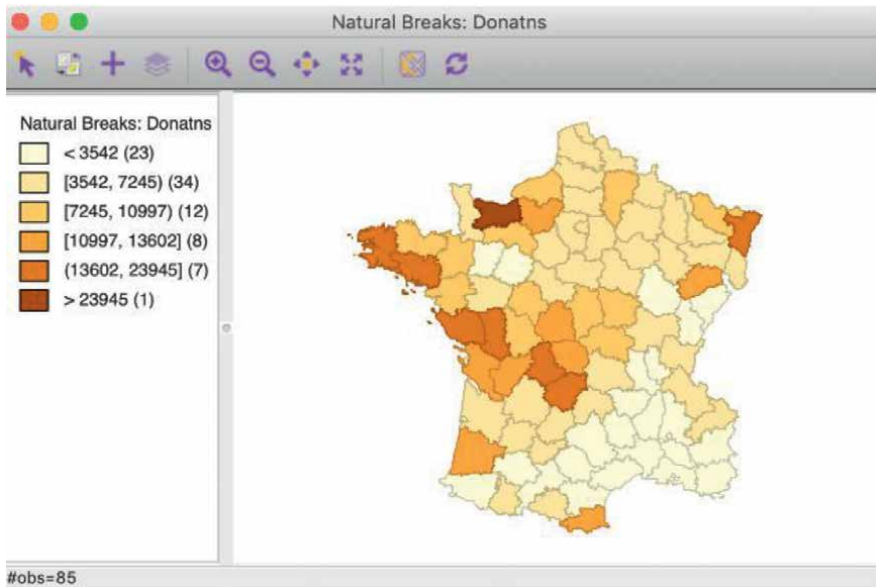


Figure 12.
Donations - natural breaks map. Source: Anselin [36].

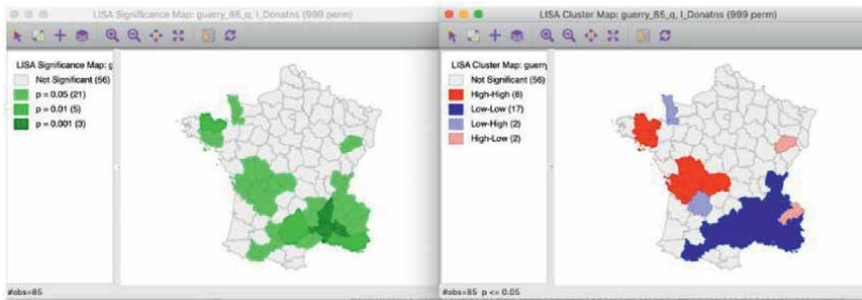


Figure 13.
LISA at a significance level of 0.05. Source: Anselin [36].

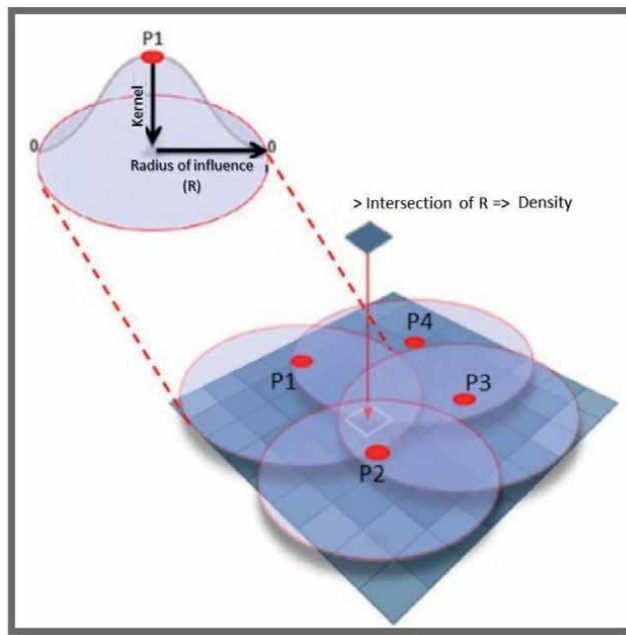


Figure 14.
Visual representation of kernel density estimation. Source: Rizzati et al. [37].

3.4 Examples of application of spatial analysis tools in health studies

There are several examples of generating geographic information at different scales of analysis in the field of health. The research by Kraemer et al. [38] projects the distribution of diseases such as chikungunya and yellow fever worldwide, considering variables such as human population, climate, and the main transmission vectors of these diseases, *Aedes albopictus* and *Aedes aegypti*, respectively (**Figure 15**).

On a more regional scale, the spatial distribution of mortality rates from heart diseases can be observed in **Figure 16**. The map depicts the rates for adults aged 35 years and older by municipality in the United States from 2018 to 2020. The concentrations of municipalities with the highest mortality rates from heart diseases, represented by the upper quintile, are primarily located in Alabama, Mississippi, Louisiana, Arkansas, Oklahoma, Georgia, Kentucky, Tennessee, and Guam. Additionally,

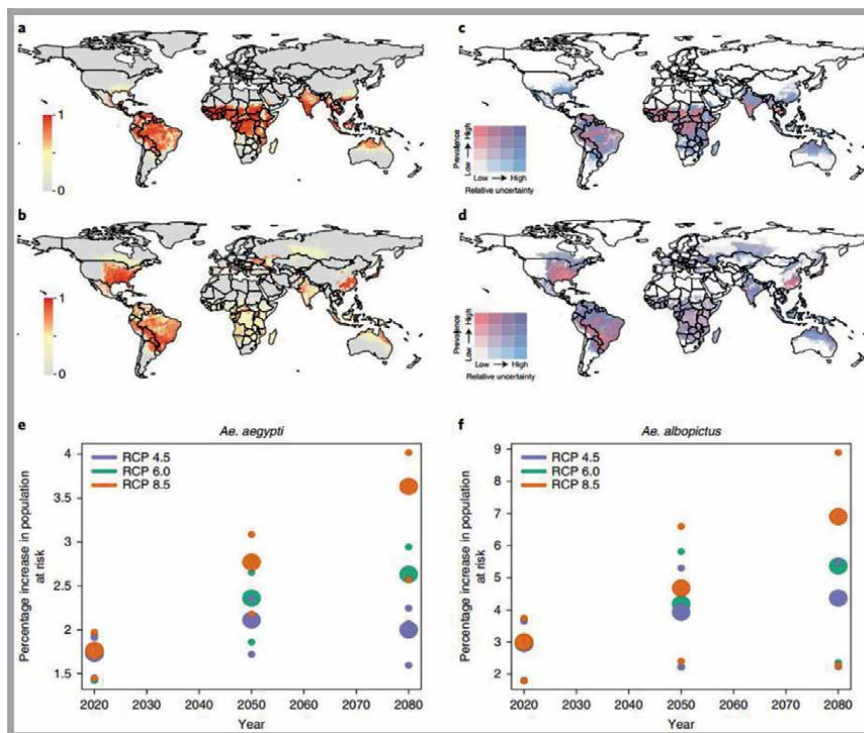


Figure 15. (a-d) Projected global geographical distribution in 2050 for *Ae. aegypti* and *Ae. albopictus* vectors. (a) *Ae. aegypti* (b) *Ae. Albopictus*. (c-d) Projection under different uncertainty scenarios (c) *Ae. aegypti* and (d) *Ae. albopictus*. Source: Kraemer et al. [38].

pockets of municipalities with high rates can be found in Michigan, Ohio, West Virginia, Virginia, North Carolina, South Carolina, Missouri, Texas, and Nevada.

An example of analysis at the local scale is the study by Tao and Cheng [39], which shows the accessibility of health services in Beijing, China (Figure 17).

3.5 The contribution of data science and machine learning in medical geography

Data science and machine learning play a crucial role in the field of medical geography, and their impact has been emphasized throughout this chapter. The intersection of data science and machine learning with medical geography represents a powerful synergy that plays a fundamental role in improving decision-making in the healthcare domain.

These disciplines provide tools and techniques that enable the analysis and interpretation of health data in spatial and temporal dimensions, thereby facilitating informed decision-making [40]. In this context, geospatial analysis, and the use of geographic information system (GIS) become fundamental allies in exploring the relationships between health and geographical location. This allows for a deeper understanding of disease patterns and the spatial and environmental determinants that influence various health conditions [41].

Medical geography greatly benefits from the application of advanced data analysis and modeling techniques [42]. The incorporation of geographic information system (GIS) and geospatial analysis allows for the investigation of complex relationships

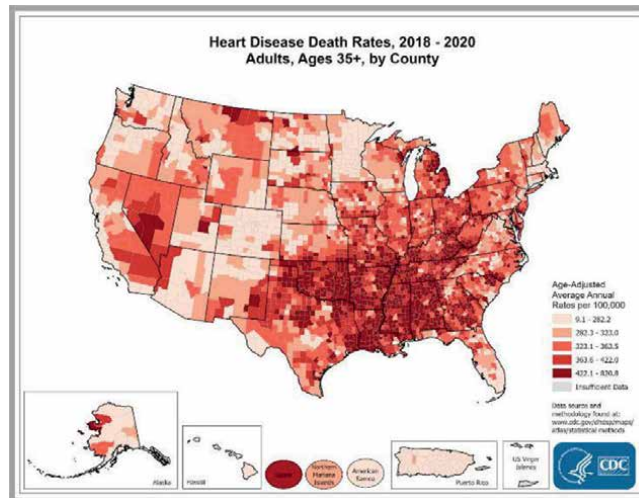


Figure 16.
Distribution of heart disease in the United States. Available via the link: https://www.cdc.gov/dhdsp/maps/national_maps/hd_all.htm.

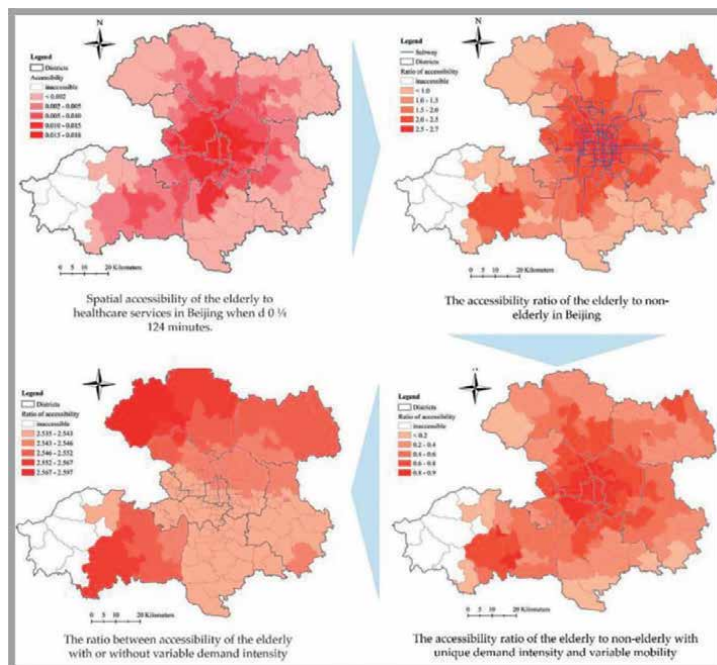


Figure 17.
Model of accessibility to health services in Beijing, China. Source: Tao and Cheng [39].

between human health and geographical location. This spatial perspective reveals disease patterns that are related to specific spatial and environmental determinants, providing a deeper insight into the epidemiology of various health conditions [41].

Machine learning algorithms are versatile tools that can be applied to large, complex medical data sets. This enables risk assessment, diagnosis, and classification, allowing

for the identification of high-risk areas and the characterization of diseases based on their impact and other relevant features [43]. The integration of data from various sources and their standardization through machine learning methods allows for the creation of unified data sets for data-driven medical research [44]. Furthermore, machine learning streamlines the integration of health data at the syntactic, structural, and semantic levels, reducing the need for manual data configuration and refinement [45].

To enhance the processing, analysis, and dissemination of health information using data science and machine learning, a simple framework is proposed to establish basic criteria for interrelating variables from various data sources.

4. Data collection

The first step is to obtain reliable data on the type of medical condition under study, tests conducted, demographic characteristics, mobility, and socioeconomic factors in the region of interest. Reliable sources include public health reports, government databases, and mobility data, i.e., data from primary and/or secondary sources.

5. Data cleaning and preprocessing

This involves addressing missing data, removing outliers, and normalizing data to make it suitable for analysis. Feature engineering tasks are also performed to create useful features from raw data.

6. Data exploration

Visualize and explore the data to identify patterns. This could include heat maps of cases, trend charts over time, and correlation analysis between variables.

7. Machine learning modeling

Utilize machine learning algorithms to predict high-risk areas. Common approaches include regression or classification models. In this case, the target variable could be the infection rate or the likelihood of an outbreak in a region.

8. Model validation and evaluation

Split the data into training and testing sets and use metrics such as accuracy, recall, and F1-score to assess model performance. Cross-validation can also be used to ensure model robustness.

9. Results interpretation

Analyze the most important features contributing to model predictions. This can help understand why certain areas are more susceptible to outbreaks.

10. Implementation and monitoring

Implement the model in a real-time system if necessary and continuously monitor data to detect changes in outbreak risk.

11. Results communication

Communicate results to public health authorities and the public. This can aid in decision-making, such as resource allocation for testing, vaccination, or social distancing measures (**Figure 18**).

These tools will have an increasingly significant impact on health geography and can assist in various ways, including identifying health data patterns and trends, predicting disease outbreaks, and optimizing the allocation of healthcare resources at different decision-making levels. When analyzing large data sets, data science, and machine learning can uncover hidden relationships and correlations that may not be evident through traditional methods, providing valuable insights into disease distribution, the identification of high-risk areas, and the specific design of interventions. Furthermore, these technologies can help monitor and track disease spread in real-time, enabling timely response and intervention strategies.

Hence, data science and machine learning have the potential to revolutionize medical geography by enhancing our understanding of health patterns and enabling evidence-based decision-making. It's essential to note that this is a simplified process, and medical geography is a complex field that may require more sophisticated models, considering factors such as population mobility, virus variants, and governmental responses. Additionally, it is crucial that the data

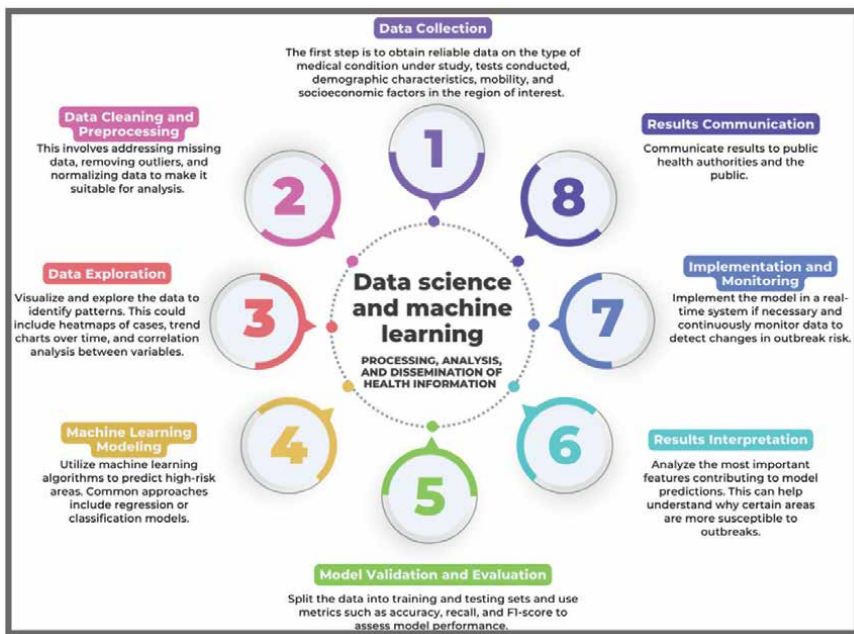


Figure 18.
Framework for enhanced health data processing and analysis. Source: The authors.

used is accurate and up to date for the predictions to be valuable in public health decision-making.

11.1 Recommendations for the application of geotechnologies in public health: toward a sustainable and equitable future

Governments should prioritize investment in the development and application of geotechnologies in their public health policies. This is based on information from various sources of first and second-order epidemiological data that can be processed using data science and machine learning. These technologies have the potential to improve resource allocation, epidemiological surveillance, and the response to health emergencies.

Healthcare sector companies should leverage geotechnologies to optimize their operations, not only in administration but also in early detection and timely assistance. This includes the use of geographic information systems for efficient management of medical facilities, route planning for the distribution of medical supplies, and spatial analysis to identify high-demand areas for healthcare services.

On the other hand, it is essential for citizens to be informed about the advantages and challenges of using geotechnologies in healthcare. This will enable them to actively participate in promoting their health and make informed decisions about access to medical services and the adoption of preventive measures.

By implementing these recommendations, we can maximize the potential of geotechnologies to enhance public health, strengthen healthcare systems, identify areas of risk, and facilitate the efficient provision of healthcare services to the population.

12. Conclusions

In conclusion, geotechnologies are a critical component in the field of public health and offer numerous opportunities to improve decision-making and resource optimization. Through the application of geographic information system, spatial analysis, and visualization tools, it is possible to gain a deeper understanding of health issues, identify spatial patterns and trends, and design effective intervention strategies.

Throughout this chapter, the diversity of applications of geotechnologies in public health has been explored, ranging from epidemiological surveillance to health services planning and environmental risk assessment. These technologies have proven to be valuable tools for the analysis and interpretation of spatial data, enabling researchers and health professionals to make informed, evidence-based decisions.

However, it is important to be aware of the challenges associated with the use of geotechnologies, such as the complexity of their implementation, associated costs, and concerns related to data privacy. Addressing these challenges through adequate training, strategic financial investment, and the implementation of robust data protection policies is critical.

Ultimately, by harnessing the potential of geotechnologies in the field of public health, we can improve the quality of life for people, optimize resource allocation, and more effectively address the health challenges facing our society. It is crucial to promote collaboration between governments, businesses, health professionals, and the community at large to maximize the benefits of these technologies and achieve a healthier and more equitable future for all.

Acknowledgements

We would like to express our heartfelt gratitude to the State University of Amapá and the Federal University of Pará for their invaluable support. We extend our sincere appreciation to Professor Marcia Aparecida da Silva Pimentel and Professor Thanan Walesza Pequeno Rodrigues for generously providing us with their time and whole-hearted cooperation in facilitating our research endeavors.

Conflict of interest

The authors declare no conflict of interest.

Notes/thanks/other declarations

This research aims to serve as a foundational resource for public servants, empowering them in their decision-making processes by providing systematically organized and geographically referenced data. These insights will be instrumental in planning activities spanning various scales, ranging from the local to the state and even national levels.

Author details

Melgris José Becerra¹ and Mariano Araujo Bernardino da Rocha^{2*}

1 Universidade Federal do Pará, Belém do Pará, Brasil

2 Universidade do Estado do Amapá, Brasil

*Address all correspondence to: mariano.rocha@ueap.edu.br

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Ness RB, Andrews EB, Gaudino JAJ, Newman AB, Soskolne CL, Stürmer T, et al. The future of epidemiology. *Academic Medicine*. 2009;**84**(11):1631
- [2] Friedman GD, Steinberg B. *Primer of Epidemiology*. New York: McGraw-Hill; 1994
- [3] Ness RB, Koopman JS, Roberts MS. Causal system modeling in chronic disease epidemiology: A proposal. *Annals of Epidemiology*. 2007;**17**(7):564-568
- [4] Halloran ME, Ferguson NM, Eubank S, Longini IM, Cummings DAT, Lewis B, et al. Modeling targeted layered containment of an influenza pandemic in the United States. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;**105**(12):4639-4644
- [5] Câmara G, Carvalho M. Análise de eventos pontuais. In: Druck S, Carvalho M, Câmara G, Monteiro A, editors. *Análise Espacial de Dados Geográficos Brasília, EMBRAPA*. 1st ed. Brasília: EMBRAPA; 2004
- [6] Gesler W. The uses of spatial analysis in medical geography: A review. *Social Science & Medicine*. 1986;**23**(10):963-973
- [7] Cromley EK, McLafferty SL. *GIS and Public Health*. New York: Guilford Press; 2011
- [8] Kirby RS, Delmelle E, Eberth JM. Advances in spatial epidemiology and geographic information systems. *Annals of Epidemiology*. 2017;**27**(1):1-9
- [9] Eberth JM, Kramer MR, Delmelle EM, Kirby RS. What is the place for space in epidemiology? *Annals of Epidemiology*. 2021;**64**:41-46
- [10] Ahasan R, Alam MS, Chakraborty T, Hossain MM. Applications of GIS and geospatial analyses in COVID-19 research: A systematic review. *F1000Research*. 2022;**9**:1379
- [11] Bowen GJ, West JB, Zhao L, Takahashi G, Miller C, Zhang T. Cyberinfrastructure for isotope analysis and modeling. *EOS. Transactions of the American Geophysical Union*. 2012;**93**(19):185-187
- [12] Xavier-da-Silva J. *Geoprocessamento para análise ambiental*. 1st ed. Rio de Janeiro: D5 Produção Gráfica; 2001
- [13] Takahashi T. *Brazil, Programa Sociedade da Informação (Brasil)*. 1st ed. Ministério da Ciência e Tecnologia: Sociedade da informação no Brasil: Livro verde, Brasília; Brasília – DF, Brasil. 2000. 203 p
- [14] Câmara G, Davis C, Monteiro A. *Introdução à Ciência da Geoinformação*. 1st ed. São José dos Campos: INPE; 2001
- [15] Fitz PR. *Geoprocessamento sem complicação*. São Paulo, SP, Brasil: Oficina de Textos; 2008. 160 p
- [16] Jensen JR. *Remote Sensing of the Environment: An Earth Resource Perspective*, Prentice Hall Series in Geographic Information Science. 2nd ed. Upper Saddle River, NJ: Pearson Prentice Hall; 2007. 592 p
- [17] de Silva AB. *Sistemas de informações geo-referenciadas: Conceitos e fundamentos*. In: *Sistemas de informações geo-referenciadas: Conceitos e fundamentos* [Internet]. 1st ed. Campinas: Unicamp; 2003 [Accessed:

- Jul 15, 2023]. pp. 236-236. Available from: <https://pesquisa.bvsalud.org/portal/resource/pt/lil-766568>
- [18] INCE F. Digital image processing systems and remote sensing†. *International Journal of Remote Sensing*. 1983;4(1):129-148
- [19] Yakamoto J, Landim P. *Geoestatística: Conceitos e Aplicações*. 1st ed. Sao Paulo, Brasil: Oficina de Textos; 2013
- [20] Druck S, Carvalho M, Câmara G, Monteiro AMV. *Análise Espacial de Dados Geográficos*. 1st ed. Brasília, DF: EMBRAPA; 2004
- [21] Tosto SG, Rodrigues C, Bolfe E, Batistella M. *Geotecnologias e geoinformação: O produtor pergunta, a Embrapa responde*. 1st ed. Brasília: EMBRAPA; 2014
- [22] Longley PA, Goodchild MF, Maguire DJ, Rhind DW, editor. *Geographical Information Systems and Science*. 2nd ed. Chichester; Hoboken, NJ: Wiley; 2005. 517 p
- [23] Gomarasca MA. *Basics of Geomatics*. 1st ed. Berlin: Springer Science & Business Media; 2009. 690 p
- [24] Magalhães MAFM, Gracie R, Barcellos C. *Sistemas de Informações Geográficas em saúde*. In: Santos SM, Barcellos C, editors. *Abordagens espaciais na saúde pública*. 1st ed. Brasília, DF: Ministerio de Saúde; 2006. pp. 43-81
- [25] Ramalho WM. *Conceitos e Arquitetura de Sistemas de Informações Geográficas*. In: Santos SM, editor. *Introdução à Estatística Espacial para a Saúde Pública*. 1st ed. Brasília, DF: Ministerio de Saúde; 2007
- [26] National Academies of Sciences, Engineering, and Medicine. *Facilitating Interdisciplinary Research* [Internet]. Washington, D.C.: National Academies Press; 2004 . Available from: <http://www.nap.edu/catalog/11153> [Accessed: Jul 15, 2023]
- [27] Stull R. *Practical Meteorology: An Algebra-Based Survey of Atmospheric Science*. 1st ed. Vancouver: UBC; 2017
- [28] IBGE. *Atlas geográfico escolar*. 7 ed. Rio de Janeiro: IBGE; 2016
- [29] Maciel DA, Diniz JMFS. *Radar: Um outro modo de mapear*. *RadarGeo*; 2021
- [30] PAHO/WHO. *Health Information Systems - PAHO/WHO*. Pan American Health Organization [Internet]. Available from: <https://www.paho.org/en/topics/health-information-systems> [Accessed: Jul 15, 2023]
- [31] Pina MF. *Armazenamento dos Dados em SIG*. In: Carvalho M, Pina MF, Santos S, editors. *RIPISA–Rede Interagencial de Informações Para a Saúde Conceitos Básicos de Sistema de Informações Geográficas e Cartografia Aplicados à Saúde*. 1st ed. Brasília, DF: OPAS-RIPSA; 2000. pp. 41-66
- [32] Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. *Google earth engine: Planetary-scale geospatial analysis for everyone*. *Remote Sensing of Environment*. 2017;202:18-27
- [33] Gesler WM, Albert DP. *How spatial analysis can be used in medical geography*. In: Albert DP, Gesler WM, Levergood B, editors. *Spatial Analysis, GIS, and Remote Sensing Applications in the Health Sciences*. 1st ed. Chelsea, Michigan: Ann Arbor Press; 2000
- [34] Rosa R. *Sistema de Informação Geográfica*. 1st ed. Uberlândia: UFB; 2004

- [35] Anselin L. Exploring Spatial Data with GeoDaTM: A Workbook. 1st ed. Center for Spatially Integrated Social Science: University of Illinois; 2005. 244 p
- [36] Anselin L. GeoDa documentation. GeoDa Workbook [Internet]. 2020 [Accessed: Jul 15, 2023]. Available from: https://geodacenter.github.io/workbook/6b_local_adv/lab6b.html#fn1
- [37] Rizzatti M, Batista NL, Spode PLC, Erthal DB, de Faria RM, Scotti AAV, et al. Mapeamento da COVID-19 por meio da densidade de Kernel. *Metodol E Aprendizado*. 2020;3:44-53
- [38] Kraemer MUG, Reiner RC, Brady OJ, Messina JP, Gilbert M, Pigott DM, et al. Past and future spread of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus*. *Nature Microbiology*. 2019;4(5):854-863
- [39] Tao Z, Cheng Y. Modelling the spatial accessibility of the elderly to healthcare services in Beijing, China. *Environment and Planning B: Urban Analytics and City Science*. 2019;46(6):1132-1147
- [40] Garg PK. Geospatial Health Data Analytics for Society 5.0. In: Garg PK et al., editors. *Geospatial Data Science in Healthcare for Society 5.0. Disruptive Technologies and Digital Transformations for Society 5.0*. Singapore: Springer; 2022. pp. 29-58
- [41] Hurd J, Hurley O, Asghari S. Issues to consider before initiating a project in medical geography. *Frontiers in Public Health*. 2016;4:1-4
- [42] Delmelle EM et al. Uncertainty in geospatial health: Challenges and opportunities ahead. *Annals of Epidemiology*. 2022;65:15-30
- [43] Boyapati S, Swarna SR, Sharma S, Agrawal R. A review of bigdata and machine learning techniques in healthcare. In: 3rd International Conference on Intelligent Sustainable Systems (ICISS). Thoothukudi, India; 2020. pp. 756-761. DOI: 10.1109/ICISS49785.2020.9315876
- [44] Tradigo G, Veltri P, Greco S. Geomedica: Managing and querying clinical data distributions on geographical database systems. *Procedia Computer Science, ICCS 2010*. 2010;1(1):979-986
- [45] Müller A et al. Machine learning for medical data integration. In: *Caring is sharing—Exploiting the value in data for health and innovation*. 2023. p. 691

Land Use and Land Cover Simulation

Hafiz Usman Ahmed Khan

Abstract

It is important to consider the dynamics of Land Use and Land Cover (LULC) patterns and how they affect society and the environment. Traditional LULC change evaluation methods can be inaccurate and require a lot of manual labor. However, AI techniques like deep learning and machine learning have shown a lot of promise for improving and automating LULC simulation processes. In this chapter, the importance of applying AI to LULC simulation is emphasized, along with the precision, efficacy, and scalability it offers. The importance of AI-based LULC simulation for making well-informed decisions in a variety of fields, including urban planning, agriculture, and natural resource management, is highlighted in the abstract's conclusion. AI-based methods have shown great promise in LULC analysis, producing accurate classification results and paving the way for subsequent simulations. These findings demonstrate the importance. These results highlight the significance of utilizing AI approaches to assist sustainable development, solve environmental issues, and influence land management choices.

Keywords: GIS, Land Use, Land Cover, simulation, AI, machine learning

1. Introduction

Cities are currently home to nearly half of the world's population, and over the next 30 years, most of the two-billion-plus person increase in global population is expected to occur in urban areas in the developing world. This represents a significant departure from the spatial distribution of population growth in the developing world over the past 30 years, which was much more evenly divided between urban and rural areas. The level of world urbanization today and the number and size of the world's largest cities are unprecedented. At the beginning of the twentieth century, just 16 cities in the world, the vast majority in advanced industrial countries-contained a million people or more. Today, almost 400 cities contain a million people or more, and about seventy percent of them are found in the developing world.

Although the terms "Land Use" and "Land Cover" are frequently used interchangeably, each term has a distinct definition. The term "Land Cover" describes the material that covers the surface of the ground, such as vegetation, urban infrastructure, water, bare soil, and so on. Land-cover identification creates the baseline data for tasks like thematic mapping and change detection analyses [1]. The term "Land Use"

describes a piece of land's function, such as agriculture, wildlife habitat, parks, industry, or recreation. When the terms "Land Use" and "Land Cover" are used together, they typically refer to the grouping or classification of human activities and natural elements on the landscape over a specific period of time using recognized scientific and statistical methods of analysis of pertinent source materials.

Land Use and Land Cover patterns must be assessed and monitored to ensure/protect natural resource management, environmental sustainability, habitat conservation, and the development of sustainable living standards for human beings. Therefore, land-cover archives must be monitored, mapped, analyzed, and updated regularly. The rate of urbanization is greater than or equal to population growth rates. Several studies emphasize large metropolitan cities, where population growth is closely related to urban land expansion based on statistical records. However, the spatial distribution, trends, patterns, and degree of urban land transition were not included in these statistics. LULC research has been a well-established methodology for computing developments on the surface of the earth, such as mapping, monitoring, modeling, and forecasting urban growth, utilizing low- to very high-resolution satellite images.

1.1 What is LULC

LULC stands for Land Use and Land Cover, which is a commonly used application in remote sensing. Remote sensing is the science of obtaining, processing, interpreting, and storing images from the ground, airborne, or space devices for many years. LULC modeling involves identifying and classifying different types of Land Uses and Land Covers based on remotely sensed images. The use of remote sensing in LULC modeling is increasing due to the availability of new sensors and multi-sensor approaches that provide insight into changes in Land Use and Land Cover.

It is a term used in the field of remote sensing and GIS (Geographic Information Systems) to describe the categorization and classification of different types of Land Use and Land Cover features on the Earth's surface [2]. Land Use refers to the human activities and purposes for which land is utilized, such as residential areas, agricultural fields, industrial zones, or commercial developments. Land Cover, on the other hand, refers to the physical and biological cover of the Earth's surface, including forests, water bodies, grasslands, urban areas, and barren lands.

LULC mapping involves the analysis and interpretation of satellite imagery or aerial photographs to identify and classify these different Land Use and land-cover categories [3]. It provides valuable information for various applications, including urban planning, environmental monitoring, natural resource management, and land-use change analysis. By understanding and mapping LULC patterns, researchers and decision-makers can assess the impacts of human activities on the environment, monitor changes over time, and develop effective strategies for sustainable land management and resource allocation.

1.2 Why do we need LULC

LULC stands for Land Use and Land Cover, which is a commonly used application in remote sensing. Land Use refers to human activity on a piece of land, while Land Cover refers to the type of coverage over the land surface, whether it is driven by human or natural forces. LULC changes are increasingly affecting human needs, and there is a growing need to map and understand these changes [4, 5]. Mapping Land Use/Land Cover (LULC) changes at regional scales are essential for a wide range of

applications, including landslide, erosion, and land planning, globally. LULC modeling involves the identification and classification of different types of Land Uses and Land Covers based on remotely sensed images. The use of remote sensing in LULC modeling is increasing due to the availability of new sensors and multi-sensor approaches that provide insight into changes in Land Use and Land Cover.

A society's social and economic development is completely correlated with its rate of expansion. This is the main justification behind socioeconomic surveys. Datasets from both spatial and nonspatial sources are used in this kind of survey. The planning, management, and monitoring of programs at the local, regional, and national levels heavily rely on LULC maps. On the one hand, this type of information aids in a better understanding of land utilization issues, and on the other, it is crucial in the establishment of the policies and programs needed for development planning. Monitoring the ongoing pattern of Land Use/Land Cover through time is essential for ensuring sustainable development. Authorities involved in urban development must create these planning models so that every available piece of land can be used in the most logical and effective way in order to ensure sustainable urban development and prevent the haphazard growth of towns and cities [6]. Information about the area's past and present land usage and Land Cover is necessary for this. We may study the changes in our ecology and surroundings with the aid of LULC maps.

We can create policies and start programs to safeguard our environment if we have inch-by-inch data on the study unit's Land Use/Land Cover.

1.3 Detection of Land Use and Land Cover change

"The process of detecting differences of state of an object or phenomenon between two different dates of the same geographical region" is the definition of LULC change. It is broken down into three sections: (a) preprocessing; (b) suitable algorithm selection for change detection; and (c) accuracy evaluation. The changes, according to remote sensing terminology, are brought on by soil moisture, climatic conditions, and spectral, geographical, thematic, and temporal restrictions. There are typically two methods for change detection: (1) post-classification comparison and (2) change detection independent of categorization. Without any prior knowledge, change detection without categorization employs the difference function map to find differences in multi-temporal remote-sensing pictures. These include change vector analysis (CVA), band stacking of different times, time series modeling (trajectory-based), and so on. In the post-classification comparison, multi-temporal images are compared for change analysis. These include composite analysis, image rationing, image regression, and aerial difference computation. Post-classification is the method that is most frequently employed. Although being sensitive to the prior categorization accuracy, it is very simple to use, provides "from-to change information," and decreases environmental differences and sensor influence. Each of these methods is used in turn to the monitoring of various LULC changes. As an illustration, consider vegetation, deforestation, monitoring of disasters, landscape fragmentation, ecological fluctuations, climate change, and urbanization. They function as a foundation for understanding the dynamics and connections between human actions and natural occurrences.

1.4 Land Use and Land Cover (LULC) classification

Classifying remote-sensing images is one of the most crucial methods for mapping LULC. Numerous categorization systems have been created and used in a variety of

contexts, including environmental change research, LULC, land resource management, spatial-temporal modeling, and change detection, to name just a few: (1) Parametric (such as the maximum likelihood classifier (MLC), minimal distance to means and the box classifier, and K-Means) and nonparametric (such as the supervised classifier, decision tree, and artificial neural network (ANN)) approaches, as well as ISODATA methods; (2) pixel- and entity-oriented; (3) automatic and semi-automatic, and (4) spectral indices. The following table describes the different classification techniques with each of the pros and cons (**Table 1**).

Decision Tree (DT)	• Nonparametric technique • Makes no conclusions on how data can be distributed. • For change and no-change classes can give a set of rules • Disadvantages: (1) Sensitive to both quality and quantity of training; (2) Data can be over-trained; (3) Does not aim for an optimum match; (4) May increase much larger in size, making it harder to analyze
Support Vector Machine (SVM)	• Nonparametric technique • Capable of managing small training data sets. • As compared to traditional methods produces high classification accuracy. • A theoretically bigger dataset can be dealt with higher dimensionality. • Outperformed RF and ANN. • In terms of average precision and robustness, the use of radial basis functions or polynomial kernels outperformed ANN and RF. • Computational involvedness is higher compared to traditional supervised methods (such as MLC) • Relatively high classification accuracy • Disadvantages: (1) Trouble in choosing the superlative kernel function; (2) With the data dimensionality, during the learning process, the computation time for classification and optimization increases polynomials.
Artificial neural network (ANN)	• Does not make assumptions about the nature of data distribution, a non – parametric technique • Estimates data properties based on training data. • As compared to conventional supervised approaches, the computational complexity is greater (such as MLC) • Disadvantages: (1) Complex architecture optimization; (2) Low computational robustness; (3) Exceeds the good mean classification accuracy due to tremendous training time; (4) The hidden layer is not well-known; (5) For network teaching, the amount of training data is essential
Random Forest (RF)	• Bagging and random algorithm based on a decision tree • Based on two parameters: (1) the number of trees, described by ‘n-tree’ and (2) in each break numerous features described by “m-try” • Much more computationally expensive than SVM classifiers • Classification trees give an individual choice of vote. • Stable and robust • Disadvantages: (1) Complexity; (2) Much laborious and time-consuming to construct several trees; (3) Overfitting; (4) No interpretability
Maximum likelihood classifier (MLC)	• Parametric approach • Involves assuming the selected signature groups in a normal distribution. • Less computational complexity • Exhibited inferior accuracies and higher variability. • Disadvantage: Requires a large training area and is unable to resolve the interclass confusion

Table 1.
Summary of the different classification techniques.

1.5 Stages in the process of image classification

As shown, the method of classifying Land Cover from satellite pictures involves several stages, including data acquisition, information data preprocessing, feature extraction, training data selection, classification, and post-processing.

1.5.1 Data acquisition

Remote-sensing data acquisition involves collecting information about the Earth's surface without physically being there. This is usually done by using instruments such as satellites, airplanes, drones, or ground-based sensors to capture images, spectra, and other data. The type of instrument used depends on the spatial and spectral resolution required for the specific application. For example, satellites are used for large-scale mapping, while drones are used for high-resolution imaging of specific areas. Data acquisition can involve active or passive sensing, with active sensors emitting energy and measuring its reflection, while passive sensors measure the energy emitted by the Earth's surface. Specifically for remote sensing image classification, data acquisition involves obtaining remotely sensed images, which can be done using satellites, drones, or other aerial platforms. Pre-processing involves preparing the images for classification by removing noise, correcting distortions, and enhancing features.

1.5.2 Preprocessing

Satellite image preprocessing is a crucial step in the process of land-use and land-cover classification. The preprocessing stage involves a series of operations that aim to enhance the quality of the satellite data and prepare it for image classification. One of the most important preprocessing steps is image correction, which involves removing distortions and errors caused by the satellite sensor, atmospheric conditions, and other factors. This is done through techniques such as radiometric and geometric correction.

Another important preprocessing step is image enhancement, which aims to improve the visual quality and interpretability of the image. This can be achieved through techniques such as contrast adjustment, histogram equalization, and filtering. Other preprocessing steps include data calibration, image registration, and image fusion. The goal of all these steps is to ensure that the satellite data is accurate, consistent, and ready for image classification, which is the next stage in the LULC classification process.

1.5.3 Image classification

Image classification of satellite data is a process used in remote sensing to interpret images captured by satellites orbiting the Earth. This process involves the identification and labeling of different features or objects in the image based on their spectral characteristics, shape, and texture. The classification of satellite data is crucial in understanding the Earth's surface and its changes over time, and it is used in various applications such as land-use and land-cover mapping, urban planning, agriculture, and environmental monitoring. The accuracy of image classification depends on several factors, including the quality of the satellite data, the choice of classification algorithm, and the expertise of the analyst performing the classification. Image classification involves the actual identification and mapping of land-use and land-cover

types using various algorithms and techniques. This can be done using supervised or unsupervised classification methods.

1.5.4 Accuracy assessment

Satellite image accuracy assessment is a crucial step in the process of land-use and land-cover classification. It involves comparing the results of image classification with a reference dataset, such as ground truth data, to evaluate the accuracy of the classification. The accuracy assessment can be performed using various statistical measures, such as overall accuracy, user's accuracy, producer's accuracy, and kappa coefficient. These measures provide information on the level of agreement between the classified image and the reference dataset. The results of the accuracy assessment can be used to improve the classification accuracy and to assess the reliability of the classification results for different applications.

Finally, accuracy assessment involves evaluating the accuracy of the classification results by comparing them to ground truth data or other reference data. This helps to identify any errors or uncertainties in the classification process and improve the overall accuracy of the results.

2. Latest trends and innovations in future simulation modeling

In recent decades, modeling of Land Use has been of increasing importance as urbanization, and LULC changes have raised concerns among decision-makers and planners about the ecosystem's and natural resources' future consequences. The increasing trend of urbanization depends on various factors, such as socioeconomic development, demography, environment, geography, and culture. This specifies the increased importance of urban areas as the focal point of the populace and commercial concentration within a particular society.

The modeling of such dynamic systems is not an easy job. Many models have been created to date, and they are divided into categories: (1) Logistic regression and Markov chain models (mathematical equation based) (2) Statically, (3) System dynamic, (4) Expert system, (5) Genetic, (6) Cellular model, (7) Hybrid or agent-based, and (8) Embedded models are all examples. The most commonly used models are (1) LCM, (2) CA, (3) Markov chain, (4) CA-Markov, (5) GEOMOD, (5) conversion of Land Use and its effects (CLUE). However, it is difficult to compare which model gives a more accurate demonstration. According to the latest research and studies, models are particularly helpful in determining the effects of urban sprawl, designing and managing LULC, and determining the correct land-use transition patterns and trends.

2.1 CLUE and GEOMOD

Veldkamp and L. O. Fresco devised the CLUE dynamic model. The study of location logistic regression is a prerequisite for this model. This model's main objective is to investigate LULC changes by including suitable driving elements including biophysical and human factors. There are three sections to this model. Regional land-use purpose, regional biophysical component, and municipal land-use distribution are the first three components. Using Costa Rica as a case study, they utilized this model to replicate changes in LULC at both the local and global levels and to illustrate how

biophysical and demographic variables have impacted LULC. Land-use conversion only takes place in situations where the new Land Use boosts yield or value due to population increase and numerous biophysical factors (including food, technology, and socioeconomic conditions). This methodology's drawback is that it requires the use of a different statistical model to predict the potential LULC.

The authors of the GEOMOD simulation model are Pontius, Cornell, and Hall. Three primary decision-making criteria are included in this model: nearest neighbors, political areas, and biophysical patterns.

2.2 Markov chain model

Markov model is a well-known and reliable model for tracking and ecological modeling, and simulating changes, trends, and predicting the future scenario at a diverse spatial scale. From one time cycle ($t = 1$) to the next ($t + 1$), this model forecasts potential LULC changes based on each LULC class's transition potential matrix. The changes are considered a stochastic process in this model. The transition matrix, on the other hand, is critical for future LULC simulations. The major drawback of this model is that it is unable to give the spatial distribution of LULC change activities.

Using the below-mentioned equation, future simulation of LULC changes can be calculated:

$$S(t + 1) = P_{ij} * S(t) \quad (1)$$

$$P_{ij} = \begin{pmatrix} P_{11} & P_{12} & P_{1n} \\ P_{21} & P_{22} & P_{2n} \\ P_{n1} & P_{n2} & P_{nn} \end{pmatrix} \quad (2)$$

$$0 \leq P_{ij} < 1 \text{ and } \sum_{j=1}^n P_{ij} = 1, (i, j = 1, 2, \dots, n) \quad (3)$$

Where $S(t)$ denotes the state of the system at time t ; $S(t + 1)$ denotes the state of the system at a time $(t + 1)$; and P_{ij} denotes the transition probability matrix from current state i to next time state j . Several research studies have been conducted to assess the future LULC scenario using the Markov chain model.

2.3 Cellular automata (CA)

CA was developed by John Von Neumann and Stanislaw Ulam to ascertain the logical underpinnings of existence. Tobler utilized the CA model for spatial modeling for the first time in 1970. This first theoretical CA technique served as the foundation for subsequent models that initially surfaced in the 1980s and had the aim of simulating and forecasting urban expansion and LULC changes [7]. The LULC and urban dynamic frameworks in the 1990s used growth and advancement in the CA model (i.e., the capacity to figure things out added to the model). CA is a dynamic technique that can describe and manage intricate and nonlinear spatial trends. It offers a clear understanding of how LULC shifts from social behavior to universal patterns. The neighborhood type, Neighborhood size, Cell size, and Transition rules make up the majority of the CA model. These settings produce the best simulation outcomes. The transition rule, which controls the model and is dependent on the training data, is

the most crucial parameter in this model. The condition of each subsequent step that depends on the current state of that cell and its surrounding neighborhood cells is referred to as the transition rule. The transition rule shows how the geographical and temporal complexity of Land Cover develops over time [8]. The ability of this model to simulate LULC and urban growth has grown. It is crucial to note that in this paradigm, space and time are separate entities. However, space is measured as a regular grid in two dimensions. This model's crucial characteristics are that they illustrate the system's complicated spatial dynamics. The CA model may be expressed using the equation below:

$$S(t, t + 1) = f(S(t), N) \quad (4)$$

Where N stands for the probability of the system's state at any time and S (t + 1) stands for the system's state at a time (t + 1). The CA model has been used in several research to evaluate changes in LULC and urbanization. These experiments show that this model is capable of accurately explaining and simulating the intricate process of LULC evolution, urban systems, and patterns [2]. This model has a key flaw in that it is unable to account for the macro-scale driving forces, such as social, economic, and cultural elements, which are in charge of LULC changes and urban sprawl. For simulation and prediction, some of this research do, however, depend on quantitative techniques like logistic regression (LR), the SLEUTH model, multi-criteria evaluation (MCE), and neural networks.

The potential of CA and Markov models to provide a computational method to give support to ecology and environmental planning, decision-making in urban, and the suitability assessment of land for construction, has been demonstrated. This is critical for the effective operation of major metropolitan areas. Many experiments, however, demonstrate the disadvantages of a single model. For LULC future prediction; integrated modeling methods are commonly used to solve the shortcomings of individual.

2.4 CA-Markov model

The CA-Markov model can address the drawbacks of a single mode by combining intricate simulation models with factual and observational models. The integrated model enhances LULC modeling, gives a better understanding, and supplements one another. In line with this, the LULC prediction model's precision varies since each research location has a distinct collection of climatic and terrain factors.

A useful tool for evaluating and modeling LULC based on existing spatial and temporal patterns is the CA-Markov model. This model may be used to simulate and recreate spatiotemporal LULC data. Markov chain and (2) are the two halves of this model. In order to provide the right LULC planning, this model may decode the Markov chain model outputs through a CA model as a spatial distribution output. The Markov model regulates the temporal transition between LULC categories based on transition matrices [7]. While taking into consideration neighborhood structure and prospective transition maps, the CA model, on the other hand, employs local rules to address changes in spatial configurations.

One of the models that is most frequently used to calculate probable LULC changes at time "t + 1" using the progression from time "t-1" to time "t" is this one. The likelihood of a change in Land Cover from one class to another during two time periods is what determines it. The probability matrices are built using historical LULC

data and anticipate future change. Numerous changes in Land Use category can be simulated, consequently, offering the ability to encourage the transition between LULC classes. Many research has recently attempted to employ the CA-Markov model to include socioeconomic and ecological variables into land-use models [9].

This will help cities better recognize and resolve a complex land-use structure and create better land-use management strategies that would better integrate urban growth and environmental protection. Compared to other models, which are also used for a similar task, this model poses benefits and drawbacks. The advantages of this model are (1) High proficiency, (2) Easy calibration, and (3) Strong capability of simulating a wide range of ground covers and dynamic patterns as compared to other models (e.g., GEOMOD and CLUE). The model's main flaws are (1) as seen in agent-based models, its failure to incorporate human, social, and economic complexities in the simulation and (2) fails to identify the new developments occurring in the studied. Therefore, to address these limitations, this model needs to be combined with other models.

For potential land-cover estimation, Soe and Le used the multi-criteria (MCE) technique in CA-Markov. The development of criteria was based on the weight assignment to the drivers of LULC changes. The more relevant the driver, the higher the weight is assigned. A criterion was divided into two categories: (1) variables and (2) restrictions. Three criteria determine it: (1) Prioritization decomposition, (2) comparative decision, and (3) Prioritization synthesis.

2.5 Land change modeler (LCM)

LCM is an incorporated model designed by Clark Labs in partnership with Conservation International to monitor and forecast LULC changes and to monitor and predict biodiversity impacts. The three parts of the LCM process—transition potential simulation, change prediction, and interpretation—are all included in the IDRISI Terrset 18.1 program [6, 10]. This model uses two thematic raster images with the same number and order of LULC groups as data. LCM evaluates LULC changes of two different periods, and provides a comparative evaluation of improvements in various LULC groups in terms of gains, losses, swaps, net changes, and overall changes, with the outcomes shown in various graphs and maps.

LCM breaks down the LULC changes for different classes, calculates and assesses their trends and patterns, and then projects these changes to predict the future LULC. Given the past transition trend, the model envisions the land-use pattern. The explanatory variables, such as distance to roads, slope, aspect, and other variables, were added to the model as raster datasets. Using Cramer's V, the influencing variables were chosen based on their availability, relative significance, and subsequent influence on LULC changes (a quantitative measure that shows the relationship between explanatory variables and land-cover classes).

The transitions are broken down into sub-models when the basic driving is taken to be the same as the ground cover transformations. For instance, the same variables that drive a change from vegetation to an urban area also cause a change from a forest to an urban region. Using MLP neural networks, the transition potential map of each sub-model is created, and explanatory variables are then allocated to each sub-model based on Cramer's V values. MLP excels at multi-transition modeling, nonlinear processing connections among variables, and converting categorical data to continuous data. The time-specific potential for change is interpreted by the transition potential maps that were developed. Future LULC maps are projected by LCM using these prospective maps and Markov chain analysis [11].

In comparison with quickly changing Land Cover, LCM produced superior forecast accuracy over a short period of time, especially in the stable Land Cover. Compared to previous models that predict LULC changes using supervised methods, such as the weighted model system, this model is more reliable. LCM developed more accurate change potential maps in which the user picks and modifies the weights. This is due to the fact that neural network outputs show changes in different LULC categories more effectively than individual probabilities obtained by weights of evidence. Furthermore, according to researchers that analyzed several approaches—including logistic regression, Bayesian analysis, weights of evidence, and a neural network—they came to the conclusion that neural networks produced predictions that were more accurate than those made by the other methods.

LCM has been utilized in several studies to forecast future LULC patterns, tropical expansion, urbanization, erosive processes, Mediterranean catchment, and habitat modeling. LCM is effective in simulating future LULC changes, patterns, and trends, according to regional case studies. This is because LCM helps local administrative authorities make decisions and gives a better grasp of LULC. All of these research findings demonstrated that LCM can provide extremely accurate LULC simulations.

2.6 Multilayer perceptron (MLP)

MLP is a feed-forward NN (Neural Network) that predicts outcomes employing the Back propagation (BP) algorithm. The core of neural net training is back-propagation (BP). It is a technique for improving the accuracy of a neural network neurons weights using the error rate from the iteration. By fine-tuning the weights, it can reduce error rates and improve the model's generalization, making it more accurate. The term "backward propagation of errors" is abbreviated as "Back Propagation." It is a popular way to train artificial neural networks. This approach is useful for calculating the gradient of a loss function with respect to all of the network's weights. It is a nonparametric algorithm that does not take multi-co-linearity into account. MLP is made up of three layers: (1) Input data, (2) secret (computing node sets), and (3) output. It signifies relationships between transitions of Land Use and their explanatory variables through a network of weighted relationships modified iteratively by the algorithm. In one direction, data flows from an input layer to an output layer *via* hidden layers, resulting in nonlinear relationships. LCM produces two kinds of predictions: soft and hard predictions. Centered on the multi-objective land allocation (MOLA) module, a hard prediction yields a forecast plot. This module assigns each pixel to one of the ground cover groups with a higher chance of being. A soft prediction represents a continuous mapping of susceptibility to change; it indicates the probability of a particular class pixel being changed to a specific pixel of Land Cover, over a T2–T1 and n-year cycle. How much land is assigned to a thematic class is determined by the transition probability matrix obtained by the Markov chain.

2.7 Artificial neural networks (ANN)

ANN is an enormously analogous distribution processor that possesses a natural storing investigational knowledge property available for use in two ways:

- Knowledge through a learning process.
- Interconnection strengths of a neuron.

A nonlinear relationship exhibited between the LULC and driving factors such as biophysical, socioeconomic, and topography such as nonlinear relationships is easily handled by ANN.

Artificial Neural Networks (ANN) are used to model complex systems such as urban sprawl, not assume the data, do not depend on exacting functional relationships, and find their way in biochemical modeling and cybernetics. ANN also reveals the relationships between future simulated urban growth probability and site attributes due to the system's nonlinear complex behavior; therefore, it perfectly fits into the regression-type model category to forecast the change in the urban Land Cover. The ANN regulates those areas that are probably going to be changed in the future, but it could not determine how much change will occur.

3. Land Use and Land Cover change modeling and future scenarios

Human activities, climate change, and natural disturbances are just a few of the many elements that can have an impact on the complicated process of land-use and land-cover change. It is essential for efficient land management and decision-making procedures to accurately simulate these changes. Remote sensing, which involves utilizing satellite or aerial photography to identify and map Land Cover, is one method of simulating changes in Land Use and Land Cover. The use of agent-based modeling, which mimics the activities and interactions of distinct agents (such as farmers or land developers) within a broader system, is another strategy [12]. It is also possible to model changes in Land Use and Land Cover using artificial neural networks and evolutionary algorithms. In these methods, patterns and correlations within the data are found using machine learning algorithms, and the linkages are then used to anticipate future changes.

It is crucial to remember that thorough validation and accuracy evaluation are necessary for accurate modeling of changes in Land Use and Land Cover. These simulations may be made more accurate by including expert comments and employing cutting-edge methods like machine learning algorithms, but continual validation is required to make sure the models are properly reflecting changes in the actual world (**Figure 1**).

3.1 Change analysis

The change between two separate time periods, time 1 and time 2, was calculated in the change analysis panel. By graphing gains and losses across various land-cover categories, the change analysis offers a fast assessment of quantitative change. Additionally, it assesses the information on Land Cover in both map and geographical representations in terms of net change, persistence, and the precise transition. These modifications are critical in locating the dominating transition from one class to another, which is subsequently categorized and targeted. The best-fit polynomial trend surface follows the pattern of change, and the spatial trend of change presents the trend as a map.

3.2 Transition potential modeling and driving forces' determination

The area of change is determined by the transition potential. If it is assumed that the underlying sources of change are the same for each transition, land-cover changes

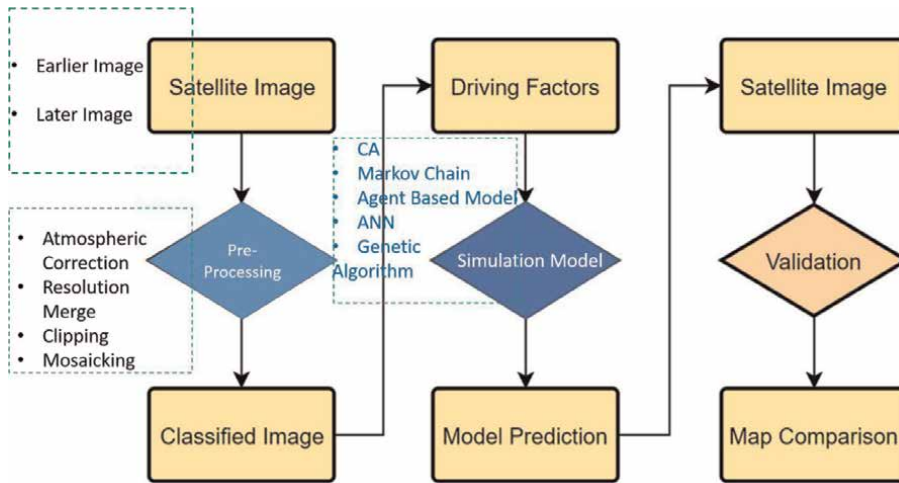


Figure 1.
Flow chart describing the process of LULC simulation by using earlier and later images.

may be categorized into sub-models. For instance, the same factors that affect the transition from forest to built-up land may also affect the transformation from agricultural to built-up land. As a result, sub-models were created for land-use and land-cover changes that shared a common set of driving factors. In order to estimate the relative frequency of various land-use and land-cover types that had taken place in the transitional areas, evidence probability was also chosen.

3.3 Selection of explanatory variables

The selection of driving factors is the most important step in modeling growth modeling that is associated with the LULC change [3, 13]. The model generates results that are close to reality. Several different factors responsible for LULC change have been identified and used by the researchers to predict the change, but still, no universal set has been formulated because each area is different from the other, and the factors responsible for change are not always common. Land transitions also vary from case to case, and the degree to which the driving factors contribute to the transitions also differs.

The urban growth prediction modeling involves computation of Cramer's V that examines the best driving variable for modeling urban growth and measures the association strength between each pair of variables used in modeling.

$$V = \sqrt{\frac{\varphi^2}{\min(k-1, r-1)}} \quad (5)$$

Where, φ^2 = Mean square contingency coefficient. k = Number of columns.
 r = Number of rows.

For the calculations, Cramer's V value falls between 0 and 1. V value ≥ 0.15 shows the influence of that particular factor, whereas > 0.4 shows a strong association between land class and factors. For modeling purposes, the driving factors need to be

declared as either static or dynamic. The factors that do not change over time are the static ones, whereas dynamic factors are time-dependent and they change with time. Dynamic variables are handled differently in the simulation process and are calculated temporally throughout prediction Multilayer Perceptron (MLP).

3.4 Change prediction

Future simulation model validation is the process of assessing the accuracy and reliability of a model that is used to simulate future scenarios. This is typically done by comparing the model's predictions with real-world data that is collected after the simulation is complete [14]. The goal of this process is to determine whether the model accurately captures the dynamics of the system being simulated and whether it can be used to make accurate predictions about the future. Validation is an important step in the development of any simulation model, as it ensures that the model is fit for purpose and can be trusted to provide reliable predictions.

3.5 Future scenario

Two different types of forecasts are generated by the Land Change Modeler: (1) hard predictions and (2) soft predictions. A projected map is generated by a hard prediction using a multi-objective land allocation (MOLA) module. Each pixel is given a land-cover class based on the likelihood that it will be present. By creating a vulnerability map and assigning a value between 0 and 1 to each pixel, soft prediction calculates the likelihood that a pixel will transition into a different land category.

3.6 Landscape matrices

Landscape matrices serve as a quantitative correlation between trends in the landscape and biological and environmental processes. They show numerical data regarding landscape structure, configuration, and dimension, allowing for comparisons over time and assisting in developing future scenarios. Landscape metrics are essential tools for defining spatial dynamics and determining how landscape elements are arranged in space and time. Broadly protected reserves, forest dynamics, natural parks, and urban extensions are all studied using them in the literature.

The metric category should represent the pattern diversity seen through the landscape, but its usage must be limited, particularly in indexes closely associated with one another. They can be classified into three categories: (1) Patch, (2) Class, and (3) Landscape. Specific patches, which reflect distinct areas with identical characteristics, are measured at the patch level. To evaluate class-level metrics, all patches of a given class type, in this case, LULC classes, are used.

In other words, it depicts the spatial distribution and pattern of a single patch from within a landscape. In simple terms, it reflects the landscape level's overall spatial pattern, taking into account all patch types at the same time. Many landscape metrics, such as the Euclidean Nearest Neighbor Distance (ENN), Largest Patch Index (LPI), and Percentage of Landscape (PLAND), can be used at the patch, class, and landscape levels [8, 15]. These metrics can be calculated at any of the three levels listed above. Complete definitions of these metrics, as well as the calculations used to calculate them, can be found in. Landscape measurements are widely used as indicators of LULC transition, ecology, water quality, and ecological health, and they are still

Category	Metric Name	Acronym Unit	Level Used	Description	Range
Patch Size and Density	Patch Density (PD)	Number of patches per 100 ha	CL	The number of patches per unit area	PD ≥ 1 , no limit
	Percentage of Landscape (PLAND)	%	CL	The aggregated area of landscape.	0–100
	Mean Patch Area (MPA)	ha	CL	An average patch size in each class	MPA > 0, no limit
Shape and Edge	Edge Density (ED)	m/ha	CL	Calculate the total lengths of all edge segments of corresponding patch type per unit area. Edge density explained the complexity of each patch shape.	ED ≥ 1 , no limit
	Largest Patch Index (LPI)	%	CL	Ratio between the largest patch of the corresponding patch type and the total landscape area.	0 < LPI $\leq$ 100
	Area Weighted Mean Fractal Dimension Index (AWMPFD)	None	LL	Measure the average fractal dimensions of patches of a particulate patch type divided by the total sum of the patch area.	1 $\leq$ AWMPFD $\leq$ 2
Proximity	Mean Euclidean Nearest Neighbor Distance (ENN_MN)	m	CL	Measure the minimum edge to edge distance to the nearest neighbor same patch type. It explains isolation of corresponding patch type or landscape.	ENN_MN > 0
Diversity and Texture	Contagion (CONTAG)	%	LL	Measure the total probability that a patch of cells neighboring the same type of cells.	0 < CONTAG $\leq$ 100
	Shannon's Diversity Index	None	LL	Indicate diversity in a landscape from the abundance of patch types. It increases as the number of different patch types increases or the distribution of area/land among patch types/classes becomes more equitable.	Shannon's Entropy ≥ 0 , no limit

Table 2.

The description of landscape metrics used for morphological analysis (where, CL = class level, LL = landscape level).

evolving. They have a suite of spatial methods for studying whole environments and the arrangement and properties of the elements that make up such landscapes. These metrics will show you how fractured landscapes are and how patches are created. They also have numerical values that can be used to describe certain classes in general, unlike segmentation. Landscape matrices may provide additional knowledge to better understand the LULC transition because of these properties (**Table 2**).

3.7 Model validation

Future simulation model validation is the process of assessing the accuracy and reliability of a model that is used to simulate future scenarios. This is typically done by

comparing the model's predictions with real-world data that is collected after the simulation is complete [16]. The goal of this process is to determine whether the model accurately captures the dynamics of the system being simulated and whether it can be used to make accurate predictions about the future. Validation is an important step in the development of any simulation model, as it ensures that the model is fit for purpose and can be trusted to provide reliable predictions.

4. Case study

4.1 Study area

The geographic coordinates of Islamabad, the capital of Pakistan, are $33^{\circ} 28'$ latitude and $73^{\circ} 22'$ longitude. According to the ICT master plan, it is organized into five zones (Figure 2). Zones I and II are separated into sections that are each around 2 km by 2 km long. Their names are assigned in the following ways: numerically from 1 to 18 from east to west, and alphabetically from A to I from north to south. Urban areas are divided into multiple equal-sized sectors by the horizontal and vertical interlacing of primary and minor highways. Included are residences, businesses, industrial centers, public and private services, foreign embassies, educational institutions, and so on. The majority of Zone-III is made up of mountains, forests, and natural landscapes.

4.2 Image classification

For this case study Landsat Images is being used and four major classes crafted details in as under (Figure 3 and Table 3).

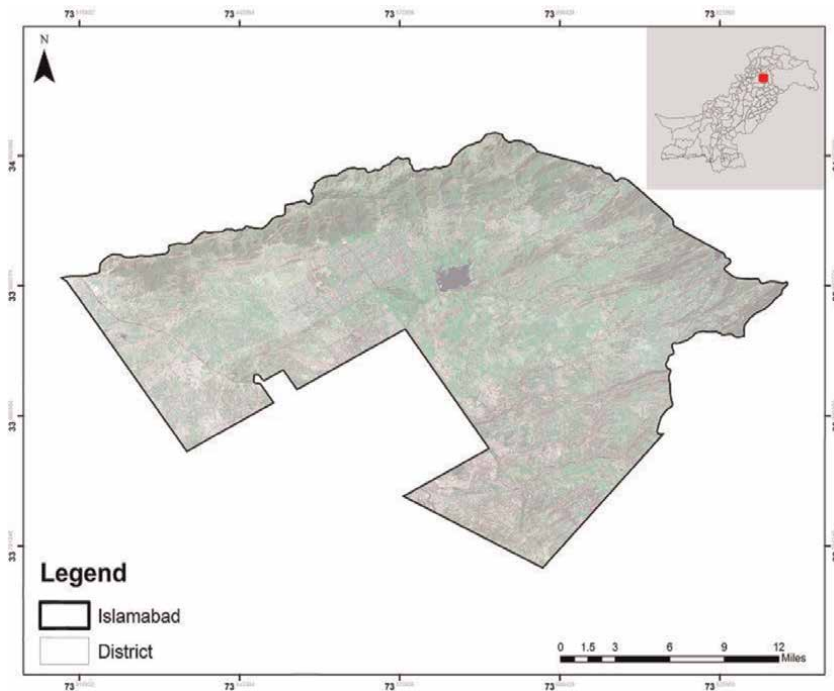


Figure 2.
Case study area Islamabad.

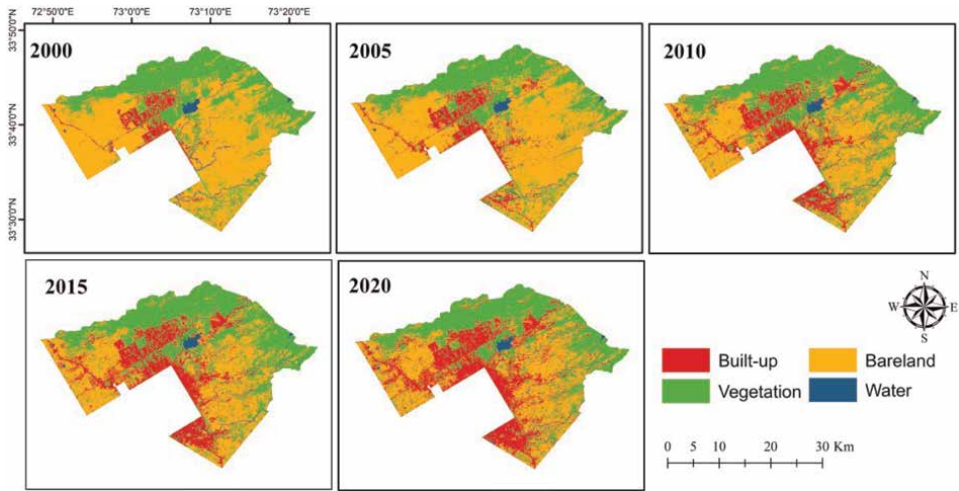


Figure 3.
Land Use Land Cover image Islamabad.

LULC Classes	Description
Bare land	Areas with no prevailing vegetation cover on at slightest 90% of the area
Built-up	Residential, industrial, and utility areas, as well as recreational areas, public installations, and utilities, are all accessible. Pavement-based roads are also included in this class.
Vegetation	Reserved woodland and fields heavily populated with trees, as well as areas where grasses dominate the landscape. Since it was difficult to distinguish between fertile fields and sparsely located villages and roads made from earthwork, they were included here.
Water	Water sources such as the Indus River and streams, as well as lakes and main rivers.

Table 3.
LULC classes description.

4.3 Major steps to conclude future simulation

The major steps in conducting a Land Use and Land Cover (LULC) future simulation using AI are as follows:

Data Acquisition: Gather relevant data sources such as satellite imagery, climate data, and socioeconomic information that include (Remote Sensing, GIS layers, and supporting datasets).

Preprocessing: Clean, align, and integrate the acquired data for analysis.

Training Data Preparation: Create a labeled training dataset by manually classifying land-cover types in the data.

Feature Extraction: Use AI techniques, such as convolutional neural networks, to extract relevant features from the training dataset.

Model Training: Train the AI model using the labeled training data to classify land-cover types accurately.

Validation and Evaluation: Validate the trained model using independent datasets to assess its performance.

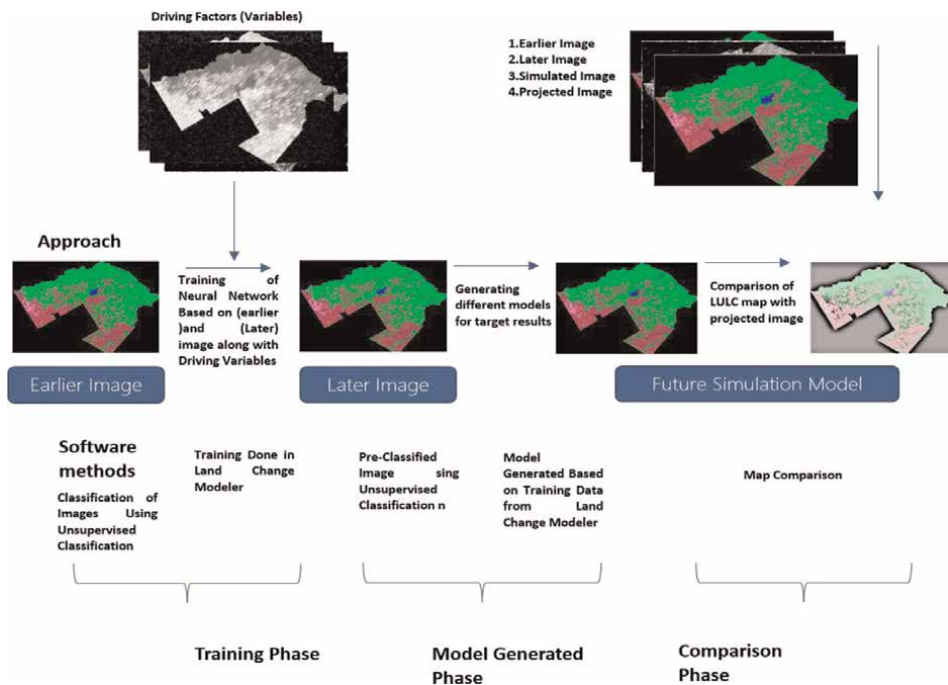


Figure 4.
 Overall workflow for LULC prediction modeling.

Future Projection and Simulation: Use the trained model to simulate future LULC scenarios based on anticipated drivers.

Uncertainty Analysis: Assess and quantify uncertainties associated with the future projections.

Scenario Analysis and Decision Support: Analyze the simulated scenarios to gain insights and support decision-making.

Documentation and Reporting: Document the methodology, findings, and limitations of the simulation study for communication purposes (**Figure 4**).

5. Why future simulation of (LULC) using remote sensing and GIS is crucial

The future simulation of Land Use and Land Cover (LULC) using remote sensing and GIS is crucial for several reasons:

5.1 Urban planning

LULC future simulation helps urban planners and policymakers to anticipate and manage urban growth and expansion. By simulating different scenarios, they can assess the potential impacts of future land-use changes on infrastructure, transportation, and the environment. This information can aid in the development of sustainable urban development plans and policies.

5.2 Environmental management

LULC simulation allows for the assessment of potential environmental changes and impacts. By projecting future land-cover changes, such as deforestation or expansion of agricultural areas, environmental managers can evaluate the consequences on biodiversity, ecosystems, and natural resources. This knowledge is essential for conservation efforts, land-use zoning, and the preservation of sensitive areas.

5.3 Climate change adaptation

With climate change, there are anticipated shifts in precipitation patterns, temperature, and other climatic factors. LULC simulation can help analyze the potential effects of climate change on Land Use and Land Cover. It aids in understanding how changes in temperature and precipitation may alter agricultural patterns, vegetation distribution, and water resources, enabling policymakers to plan for adaptation strategies.

5.4 Disaster risk assessment

Future simulation of LULC can assist in assessing vulnerability and risk to natural disasters. By analyzing land-cover changes and their potential impacts, emergency management agencies can identify areas at high risk of flooding, landslides, or other hazards. This knowledge supports the development of early warning systems, land-use regulations, and disaster preparedness plans.

5.5 Policy and decision-making

LULC simulation provides valuable information for policymakers and decision-makers. It enables them to evaluate the effectiveness of current land-use policies, anticipate future challenges, and design appropriate strategies. By incorporating spatially explicit future scenarios, decision-makers can make informed choices to promote sustainable development, resource management, and socioeconomic growth.

6. Factor to consider for LULC simulation

Choosing data for LULC (Land Use and Land Cover) simulation involves considering several factors to ensure the accuracy and reliability of the simulation results. Here are some steps to guide you in selecting appropriate data:

6.1 Define the study area

Determine the geographical extent of your study area. Consider the size, location, and specific characteristics of the area to ensure the data you choose aligns with the study objectives.

Identify Data Requirements: Clearly define the data requirements for your LULC simulation. This may include spatial resolution, temporal coverage, thematic detail, and specific data formats. Consider the scale and level of detail required for your analysis.

6.2 Acquire Base data

Start by obtaining base data that provides foundational information about the study area. This may include digital elevation models (DEM), land-cover maps, topographic maps, satellite imagery, and existing GIS datasets. Sources for these data can include government agencies, research organizations, open data portals, or commercial providers.

6.3 Assess data availability

Evaluate the availability and accessibility of the data sources that meet your requirements. Check if the data is freely available, requires purchase or licensing, or if you need to request it from relevant organizations. Consider the data format, compatibility with your GIS software, and any processing or preprocessing steps required.

6.4 Data compatibility and consistency

Ensure that the data sources you choose are compatible with each other in terms of spatial reference system, coordinate system, and projection. Consistency across different data layers is crucial to maintain accuracy and avoid discrepancies in the simulation.

6.5 Data quality assessment

Assess the quality and reliability of the data sources. Consider factors such as data accuracy, resolution, completeness, and currency. Look for metadata and documentation that describe the data collection methods, validation procedures, and potential limitations.

6.6 Data integration and fusion

Depending on your simulation requirements, you may need to integrate multiple datasets to capture different land-use and land-cover categories accurately. Data fusion techniques, such as merging satellite imagery with ground-based data or combining different data sources, can improve the accuracy and representativeness of your simulation [17].

6.7 Validation and ground Truthing

Plan for validation and ground-truthing activities to verify the accuracy of your selected data. Field surveys, ground-based measurements, or comparison with existing ground truth data can help validate the simulated LULC outputs and refine the accuracy of the model [11, 18].

7. Conclusion

In summary, LULC future simulation using remote sensing and GIS is essential for informed decision-making, environmental management, urban planning, climate change adaptation, and disaster risk assessment. It helps us understand and anticipate

the potential impacts of future land-use changes, facilitating sustainable and resilient development. Remember, the choice of data for LULC simulation should align with the specific objectives and requirements of your study. It is crucial to consider data quality, compatibility, availability, and validation to ensure reliable and accurate simulation results.

Acknowledgements

The authors would like to thank USGS Earth Explorer for providing Landsat data. The authors would also like to thank the anonymous reviewers for their insightful comments and substantial help in improving this chapter.

Conflicts of interest

The authors declare no conflict of interest.

Data availability statement

The data presented in this study are available on request from the corresponding author.

Author details

Hafiz Usman Ahmed Khan
Department of Space Technology Application, Beihang University, Beijing, China

*Address all correspondence to: hafiz_usmankhan@hotmail.com

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Khawaldah HA, Farhan I, Alzboun NM. Simulation and prediction of land use and land cover change using GIS, remote sensing and CA-Markov model. *Global Journal of Environmental Science and Management*. 2020;**6**(2):215-232
- [2] Saddique N, Mahmood T, Bernhofer C. Quantifying the impacts of land use/land cover change on the water balance in the afforested River Basin, Pakistan. *Environmental Earth Sciences*. 2020;**79**(19):1-13
- [3] Wang R, Murayama Y. Change of land use/cover in Tianjin City based on the Markov and cellular automata models. *ISPRS International Journal of Geo-Information*. 2017;**6**(5):150
- [4] Congalton RG, Gu J, Yadav K, Thenkabail P, Ozdogan M. Global land cover mapping: A review and uncertainty analysis. *Remote Sensing*. 2014;**6**(12):12070-12093
- [5] Mannan A, Liu J, Zhongke F, Khan TU, Saeed S, Mukete B, et al. Application of land-use/land cover changes in monitoring and projecting forest biomass carbon loss in Pakistan. *Global Ecology and Conservation*. 2019;**17**(January):e00535
- [6] Liu Y, Hu Y, Long S, Liu L, Liu X. Analysis of the effectiveness of urban land-use-change models based on the measurement of spatio-temporal, dynamic urban growth: A cellular automata case study. *Sustainability (Switzerland)*. 2017;**9**(5):1-15
- [7] Kityuttachai K, Tripathi NK, Tipdecho T, Shrestha R. CA-Markov analysis of constrained coastal urban growth modeling: Hua hin Seaside City, Thailand. *Sustainability (Switzerland)*. 2013;**5**(4):1480-1500
- [8] Meng X, Zhang M, Wen J, Du S, Xu H, Wang L, et al. A simple GIS-based model for urban rainstorm inundation simulation. *Sustainability (Switzerland)*. 2019;**11**(10):2830
- [9] Dadashpoor H, Malekzadeh N. Driving factors of formation, development, and change of spatial structure in metropolitan areas: A systematic review. *Journal of Urban Management*. 2020;**9**(3):286-297
- [10] Ramachandra TV, Bharath AH, Sowmyashree MV. Monitoring urbanization and its implications in a mega city from space: Spatiotemporal patterns and its indicators. *Journal of Environmental Management*. 2014;**148**: 67-81
- [11] Gilani H, Ahmad S, Qazi WA, Abubakar SM, Khalid M. Monitoring of urban landscape ecology dynamics of Islamabad capital territory (ICT), Pakistan, over four decades (1976–2016). *Land*. 2020;**9**:123. DOI: 10.3390/land9040123
- [12] Hasan L. The Islamabad Master Plan. 2020
- [13] Bokhari SA, Saqib Z, Ali A, Haq MZ. Perception of residents about urban vegetation: A comparative study of planned versus semi-planned cities of Islamabad and Rawalpindi, Pakistan. *Journal of Ecosystem & Ecography*. 2018;**8**(1):1-8
- [14] Araya YH, Cabral P. Analysis and modeling of urban land cover change in Setúbal and Sesimbra, Portugal. *Remote Sensing*. 2010;**2**(6):1549-1563
- [15] Dewan AM, Yamaguchi Y. Using remote sensing and GIS to detect and monitor land use and land cover change in Dhaka metropolitan of Bangladesh

during 1960-2005. Environmental Monitoring and Assessment. 2009; **150**(1–4):237-249

[16] Jia Y, Tang L, Xu M, Yang X. Landscape pattern indices for evaluating urban spatial morphology – A case study of Chinese cities. Ecological Indicators. 2019;**99**(November 2018):27-37

[17] Salih AAM, Ganawa ET, Elmahl AA. Spectral mixture analysis (SMA) and change vector analysis (CVA) methods for monitoring and mapping land degradation/desertification in arid and semiarid areas (Sudan), using Landsat imagery. Egyptian Journal of Remote Sensing and Space Science. 2017;**20**:S21-S29

[18] Khan MS, Ullah S, Sun T, Rehman AU, Chen L. Land-use/land-cover changes and its contribution to urban heat island: A case study of Islamabad, Pakistan. Sustainability (Switzerland). 2020;**12**(9)



Edited by Rifaat Abdalla

Dive into the dynamic world of Geographic Information Systems (GIS) and data science with our comprehensive book in which innovation and insights converge. This book presents a pioneering exploration at the intersection of GIS and data science, providing a comprehensive view of their symbiotic relationship and transformative potential. It encapsulates advanced methodologies, real-world applications, and interdisciplinary approaches that redefine how we perceive and utilize spatial data. Offering a gateway to cutting-edge research and practical insights, this book serves as a crucial resource for scholars, practitioners, and enthusiasts alike. It addresses pressing challenges across diverse domains, from environmental studies to public health and predictive analytics, demonstrating the paramount significance of integrating GIS with data science methodologies. It is an essential compass guiding readers toward a deeper understanding and application of these dynamic fields in today's data-driven world.

Published in London, UK

© 2024 IntechOpen
© Ungrim / iStock

IntechOpen

