Earth Science Informatics
https://doi.org/10.1007/s12145-019-00378-z

RESEARCH ARTICLE

# Online image classification and analysis using OGC web processing service

Hariom Singh[1] · R. D. Garg[1] · Harish C. Karnatak[2]

Received: 15 June 2018 / Accepted: 2 January 2019


## Abstract

The online geoprocessing and analysis using state-of-the-art technology are offering automated analytical tools to a large group of users. This paper describes an online geoprocessing and analysis framework, which allows an inexperienced user to perform unsupervised classification of remote sensing data. The parallel computing based geoprocessing and analysis framework adopted in this work has been implemented using Free and Open Source Software for Geospatial (FOSS4G). Web Processing Service (WPS) based geoprocessing framework facilitates the deployment of unsupervised classification algorithm on the web in a standardized way. However, it is dynamic in nature to deploy other geoprocessing algorithms. The developed system describes how to process remote sensing data (Sentinel-2, Landsat-8, etc.) for classification and sharing the interoperable results in a distributed environment. To validate the classification results, a prototype architecture based on participatory GIS is developed for field data collection, accuracy assessment and online dissemination. The accuracy assessment (i.e., overall accuracy, Kappa coefficient) is performed to validate the derived classification results using collected field data.

**Keywords** Distributed GIS · Parallel computing · Unsupervised classification · Web processing service · Participatory GIS

## Introduction

The web GIS based portals provide uniform and centralized interfaces to access the heterogeneous and distributed data archives (Karnatak et al. 2012; Sun et al. 2016). The different types of web GIS portals are available for geovisualization and querying for specific theme and are targeted to specific class of users (Rathore et al. 2010; Porta et al. 2013; Dhamaniya et al. 2016; Singh and Garg 2016). The developed geoportals do not satisfy the requirements of geoprocessing and analysis in the distributed environment. The number of distributed GIS portals focusing on geoprocessing and analysis are very less. At present, distributed GIS-based systems are very much required to facilitate online geoprocessing and analysis for various geoscientific applications. The Open Geospatial Consortium (OGC) has published various Service Oriented Architecture (SOA) based geoprocessing standards for data dissemination, geovisualization, geoprocessing, etc. (Li et al. 2011). The exchange and collaboration of geospatial data are performed by OGC web services such as Web Map Service (WMS), Web Coverage Service (WCS), Web Feature Service (WFS), Web Processing Service (WPS), etc. (Castronova et al. 2013). The interoperability of network-enabled geoprocessing and analysis is carried out with the utilization of OGC based WPS standard. It provides a web-based method for all kind of geoprocessing and analysis including any algorithm or geocomputation model that operates on georeferenced data. It allows organizations to deliver geoprocessing and analysis operations as well as web processing model to users independent from the underlying software (Michaelis and Ames 2009; Giuliani et al. 2012; Bastin et al. 2013; Singh et al. 2017). The implementation of online geoprocessing is performed with the WPS framework such as 52° North WPS, ESRI ArcGIS Server, PyWPS, etc. The WPS frameworks depend on the different development environment and have their pros and cons. Karnatak et al. (2017) described

Communicated by: H. A. Babaie

✉ R. D. Garg
garg_fce@iitr.ac.in; rdgarg@gmail.com

1 Geomatics Engineering Group, Department of Civil Engineering, Indian Institute of Technology Roorkee, Roorkee, India

2 Indian Institute of Remote Sensing, ISRO, Dehradun, India

Published online: 02 February 2019

in detail about WPS based geoprocessing architecture at the conceptual level.

At present, mostly unsupervised classification methods on remote sensing data are performed on desktop-based systems (Flanders et al. 2003; Long and Srihann 2004; Sun et al. 2016). The unsupervised classification, an efficient approach for information extraction on images has already been performed as image analysis software tool in Commercial Off-The Shelf Software (COTS) based desktop software system, e.g., ArcGIS, ENVI, ERDAS, etc. The critical disadvantage of this type of software is operational cost while upgrading previous versions and performing specific geoprocessing operations. Most of them are platform dependent and require technical skills for performing geoprocessing operations. The distributed GIS-based geoprocessing systems, which can be accessed from anywhere with the Internet connection, are platform-independent and allow the domain expert to focus on geoprocessing algorithms instead of software tools. These software offer many advantages over COTS-based desktop software system such as accessibility, interoperability, support for real-time spatial analysis, sharing geoprocessing services, etc. (Dubois et al. 2013; Karnatak et al. 2017). Hence, the development of online geoprocessing system has been identified as an important research paradigm to facilitate remote sensing data classification for common users. Furthermore, the demand of parallel computation based on satellite data is constantly increasing with the rise of big geospatial data for application of geoprocessing and analysis (Qin et al. 2014; Astsatryan et al. 2015). In high performance and parallel computing approach, the geospatial dataset is divided into smaller objects and one object is assigned to each processor, each object is computed independently of the others. After that, it combines the derived results of several geocomputation into the final output (Giuliani et al. 2012; Singh et al. 2017). Furthermore, the sharing and dissemination of geospatial data on the web have triggered a great impulse. After the evaluation of Web 2.0 (O'Reilly 2007), Internet and Communication Technology (ICT) evolved from traditional desktop-based model to a collaborative model, where the user can interactively create and share information contents (Brovelli et al. 2008; Brovelli et al. 2015). The OGC standards for interoperability, i.e., WMS, WFS, WCS, WPS, etc., and state-of-the-art technological development have revolutionized the field of geospatial domain known as GeoWeb (Brovelli et al. 2008; Haklay et al. 2014; Alonso et al. 2017). The GeoWeb has developed very fast in the last decades and it is now commonplace for geospatial data and web application. It has opened up a new dimension to distributed GIS application (e.g., mashups, volunteered geographic information or participatory GIS). As a consequence, the term GeoWeb 2.0 is the new era of interactive and dynamic tools for allowing user participation in managing geospatial data. Another remarkable boost to this new era of GeoWeb 2.0 is given by the mobile devices provided with sensors (e.g., GPS receivers and cameras) which are able to collect geospatial data and publish it in the interoperable environment. The paradigm of participatory GIS was initiated in the mid-1990s to describe the use of GIS as a tool for public participation in decision-making processes (Haklay et al. 2014). In the same way, it is described as an incredibly fruitful platform in the scientific developments of GeoWeb 2.0. As the requirements of the geoscientific application for group decision making and validation, the development of participatory GIS tool is required to visualize geospatial information collected in situ from the mobile device (Maisonneuve et al. 2010). The received data through participatory GIS can be used for many applications, including Land Use Land Cover (LULC) map validation.

This paper has proposed distributed GIS-based system framework to demonstrate parallel unsupervised classification using remote sensing images. The framework is generic and is based on Free and Open Source Software for Geospatial (FOSS4G). It is developed by using a processing backend of R programming language and statistical computing environment using 52° North Web Processing Service (WPS4R) framework (R Development Core Team 2016). The developed framework comprises an efficient geoprocessing engine to deploy and execute unsupervised classification algorithm. In particular, this is an exercise on one classification algorithm, but the geoprocessing methodology used in this study can be replicated for other geoscientific applications. Hence, it is robust and scalable to deploy any algorithm and geocomputation Here, the crucial steps are demonstrated using different modules, especially geoprocessing engine. An annotated R-script of unsupervised classification algorithm using random forest is deployed under the geoprocessing framework. The parallel computing based on classification algorithm is performed to improve the efficiency and capability of geoprocessing engine.

Initially, the geospatial datasets are published as web services (WFS, WCS) for data access and interoperability. Thereafter, the pre-and post-processing operations are automatically performed by the developed system for reducing the time and computational resources. Here, the distributed GIS-based parallel unsupervised classification algorithm is implemented and demonstrated for geoprocessing and analysis using remote sensing images (Sentinel-2 and Landsat-8) of the Haridwar district of India. To access the classification results, Graphical User Interface (GUI) of online geoprocessing environment is implemented for geovisualization, automation in raster symbology based on the equal interval, downloading, etc. The developed online software offers the ways for sharing and collaboration the geoprocessing functionality with other distributed GIS clients. To validate the derived results of classification, participatory GIS tool is developed for field data collection, accuracy assessment and interoperability. The collected field data is used as reference classes for LULC map validation. The accuracy assessment is performed with the collected reference classes and derived classification results (Landsat-8, Sentinel-2) of the online geoprocessing and analysis. The evaluation of classification results is carried out with the accuracy assessment parameters such as producer's accuracy, user's accuracy, overall accuracy and Kappa coefficient. The experimental work demonstrates the theoretical and practical aspect for the future development and demonstration of an online unsupervised classification system in an operational environment.

## Related work

The Spatial Data Infrastructure (SDI) is a framework of geospatial data, metadata and tools that are connected to utilize spatial data in flexible and efficient way (Rautenbach et al. 2013; Yue et al. 2016). However, additional geoprocessing services on SDI is required for web-based computation of heterogeneous and distributed spatial data archives (Lanig and Zipf 2009). The WPS standard based on the SOA paradigm facilitates to build an online and interoperable geoprocessing environment. The various case studies have used WPS standard to process remote sensing data for different applications (Michaelis and Ames 2009; Giuliani et al. 2012; Castronova et al. 2013; Hinz et al. 2013; Karnatak et al. 2017). In the same way, Chen et al. (2012) have enabled WPS based geoprocessing framework for remote sensing data processing and analysis. This platform consists of GUI and a WPS layer for geoprocessing. Another case study by Giuliani et al. (2012) represented processing of distributed geospatial data using OGC WPS specification. The concept of WPS is also presented to develop, eHabitat, a computational environment for ecological modelling (Dubois et al. 2013). An interoperable application for geoprocessing of remote sensing images with the index, i.e., Green Atmospherically Resistant Index (GARI), Normalized Difference Vegetation Index (NDVI) has been developed with the utilization of Python programming based geoprocessing framework (PyWPS) (Astsatryan et al. 2015). PyWPS is written with support for GRASS GIS, allowing from GRASS modules to run as geoprocessing programs. The WPS specification support python programming language and environment. It is implemented using Common Gateway Interface (CGI) based scripting environment. Apart from this, an online spatial biodiversity model has been developed as case study of OGC WPS standard (Singh et al. 2017). However, very few studies are available for the analysis of unsupervised classification in online geoprocessing using OGC WPS standard.

## Unsupervised classification

Unsupervised classification is a type of pixel-based automated classification approach. The user specifies the number of classes, which depend on the numerical information in the datasets (Smits et al. 1999; Maulik and Sarkar 2012; Quirita et al. 2016). The clustering algorithm is used to group the image pixels based on their spectral similarity. The computation procedure of unsupervised classification algorithm is commonly automated. However, the users need to control over specific inputs. It comprises the maximum iterations, (how many times the classification algorithm runs), number of classes, etc. After the image data classification, the user needs to interpret and analyze, merge classes and provide symbology accordingly. The cluster classification approach is relatively quick and easy to run (Zehtabian and Ghassemian 2017). In this approach, the classes are generated purely based on spectral information and it is not subjective as manual visual interpretation. The schematic representation of the unsupervised classification using activity diagram is described in Fig. 1. Each band of satellite imagery (Sentinel-2, Landsat) as input tiles is published as data

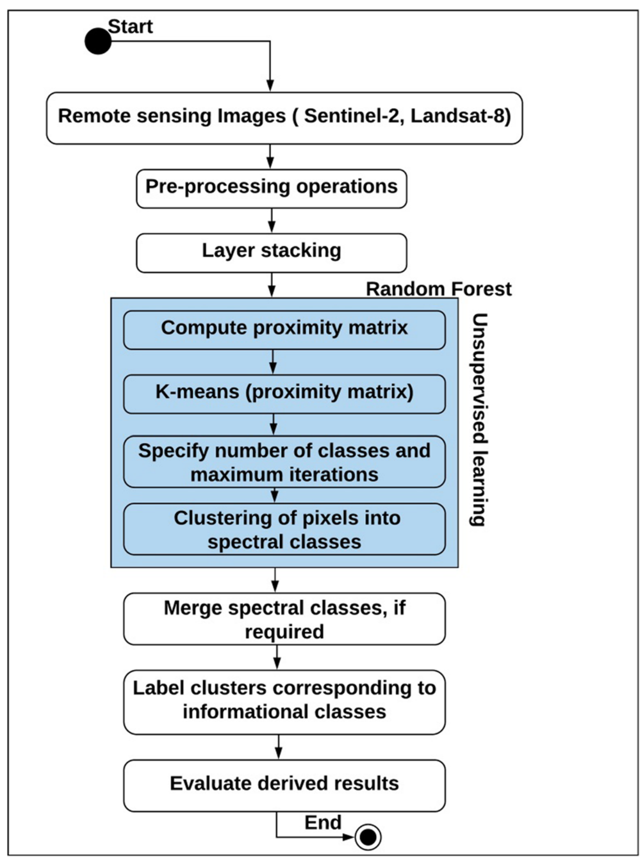


**Fig. 1** Activity diagram of unsupervised classification approach

services in GIS server. Also, the pre-processing operations such as crop/mask and downloading AOI are performed and resampling operation is carried out to similar scale and resolution. Thereafter, layer stacking is performed for random forest-based unsupervised clustering. Merge spectral classes, if required and level cluster corresponding to LULC classes. The classes are automatically generated using assigned raster symbology of the derived results as the post-processing GIS operations of the geoscientific application workflow. Further, there are many algorithms for unsupervised classification (random forest, K-means, etc.); each algorithm can cluster pixels into spectral classes for a better outcome. Also, subjective human intervention is the key method to label clusters corresponding to information classes. The optimal classification results generally require the human observation and fine-tuning approach. Based on the procedures of the pixel-based unsupervised classification, this study proposes an online geoprocessing and analysis framework for unsupervised classification in next section.

## Materials and methods

In the present study, the distributed GIS-based framework is developed which comprises five modules such as input data, parallel processing, geoprocessing engine, GUI and services. The five modules mutually communicate effectively with ICT. Here, the input data module is responsible for storing the datasets in the database server and publishing the data as OGC web services via geospatial data server (local server or remote server). The parallel processing is used to facilitate the algorithms quickly, efficiently and reliably. The geoprocessing engine provides an environment to execute the algorithm and geocomputation model on the web in a standardized way. The GUI module is responsible to request/response operations with users and sends their request to the geoprocessing module and return a response to the user's browser. The service module provides the data and geoprocessing services to the other clients. The proposed system architecture for standardized web-based geoprocessing service of the unsupervised classification algorithm is given in Fig. 2. This study provides

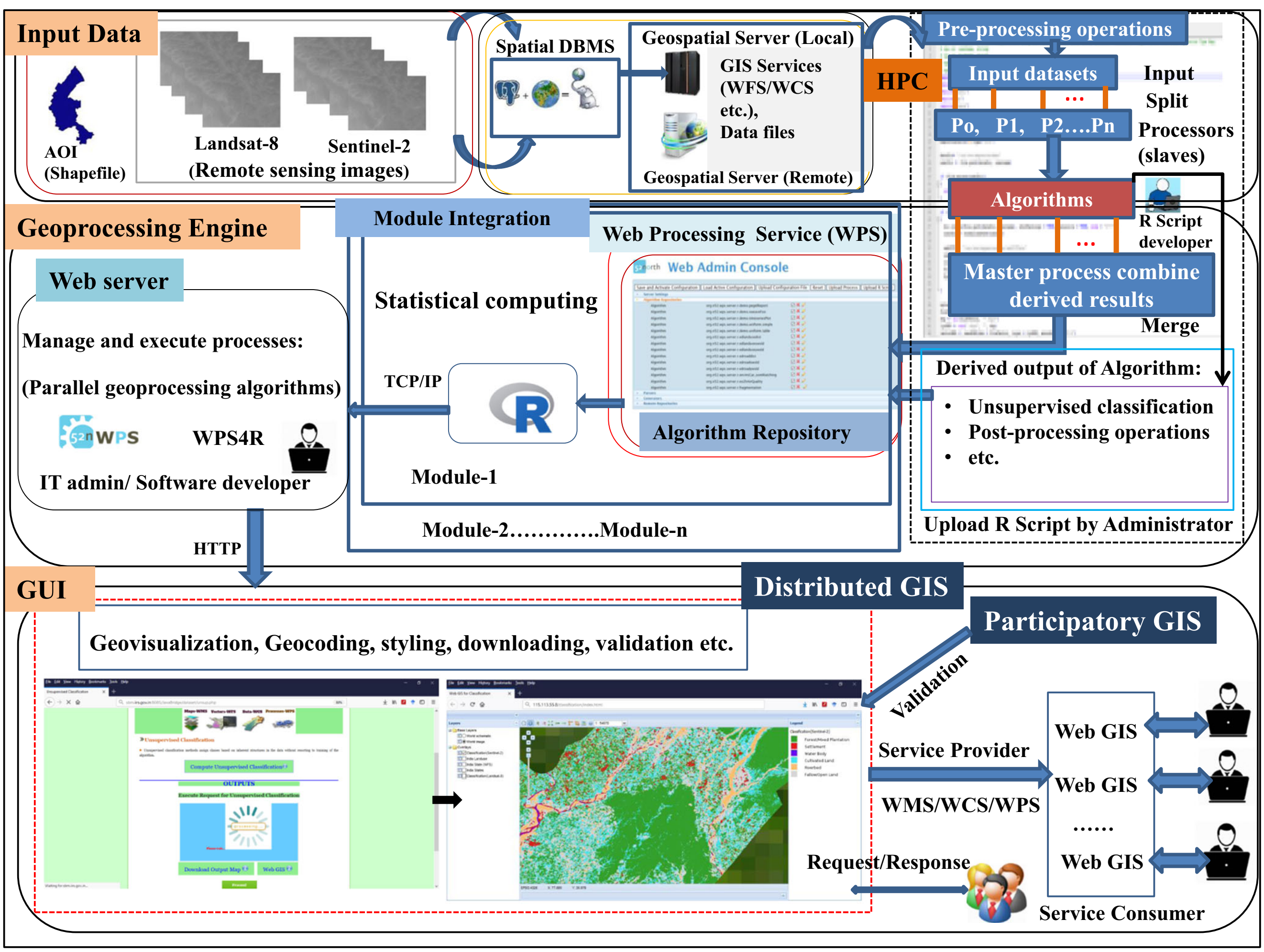


Fig. 2 System architecture for online image classification and analysis using WPS

a generic approach to process algorithms in the parallel computing environment, which is based on the linkages of R statistical computing and OGC WPS using 52° North WPS4R framework. To achieve the unsupervised clustering, the developed online application is effective in all the aspects of geoprocessing and analysis in a single integrated environment, i.e., interoperability, scalability, high performance and parallel computing, reusability, flexibility, distributed computing and real-time field data collection using participatory GIS approach. The detailed description of each module is given in the following section.

In this study, the remote sensing images (Sentinel-2 and Landsat-8) of the year of 2017 are evaluated for the unsupervised classification. In order to perform data dissemination and interoperability, Sentinel-2 data bands (2–8, 11–12) and Landsat-8 data bands (1–7, 9) are considered and published as WCS using geospatial data server, e.g., GeoServer (http://geoserver.org/). The vector data of the Area of Interest (AOI) is also published as WFS to determine the extent of the study area. Therefore, the developed geoprocessing framework facilitates to integrate input datasets through distributed geospatial information service archives using local or remote geospatial server. Besides, the user can also upload the remote sensing datasets into the application server repository from the local file system. Furthermore, the pre-processing operations (i.e., crop/mask, downloading, resampling, etc.) are carried out to prepare input datasets for the unsupervised classification algorithm. Here, the concept of parallel programming paradigm is used to enhance the performance of the algorithms. The purpose of parallelizing the algorithms is to reduce the time needed to derive the output while the quality of the output is the same as in the algorithm which is sequential. In order to obtain parallel programming in proposed algorithm, the input datasets are divided into multiple-core and then the cores (processes) compute the results independently. After that, it merges the derived results into final result from all the computing process. The advantage of this approach is to enhance the computational efficiency to handle large-scale geospatial datasets. Here, R programming language and statistical computing environment are used to develop geoprocessing engine for high performance and parallel processing. The geoprocessing engine is responsible for online geoprocessing and analysis in the parallel computing environment. The 52° North WPS4R enables the R programming backend and deployment environment for geoprocessing on the web in a standardized way. The framework considers the annotated R-script of the experimental algorithms with the post-processing operations, which are automatically performed in order to obtain classified results. Once the process is complete, the derived results can be stored as accessible web resources (WMS, WCS and WPS) or can be downloaded as classified raster layer, which is generated through automatic classification approach to facilitate raster symbology.

As an experiment of online geoprocessing, the unsupervised algorithm is deployed by the administrator as module-1. Other algorithms can be deployed and integrated as modules-2, module-n, if any. Here, TCP/IP protocol allows binary request/response to be sent to R statistical computing environment. The package Rserve supports remote connection, file transfer and user authentication. The geoprocessing framework is deployed under the execution environment of web server, i.e., Apache tomcat server. The web server manages sessions and provides an execution environment across the high-speed network. Similar to the desktop-based unsupervised classification systems, GUI of web-based unsupervised classification system essentially includes request/response environment. At the end, the module, service is responsible for sharing data service (WMS, WFS and WCS) and geoprocessing service (WPS of unsupervised classification) to the other distributed clients. It exploits the power of distributed geoprocessing of geospatial data located anywhere on the Internet and allows organizations to reuse the same geoprocess (algorithms) for multiple applications. The accessibility of geoprocess can be restricted using a variety of security mechanisms such as HTTPS and firewalls. In order to qualitatively evaluate the classification results, the distributed GIS-based system is developed by using FOSS4G such as HTML5, PHP, OpenLayers, GeoExt, ExtJS, MapServer, GeoServer and PostgreSQL along with PostGIS database server.

Furthermore, the participatory GIS-based prototype is developed for validation data collection on the field in the real-time environment. The detailed description of system architecture of participatory GIS is shown in Fig. 3. The developed prototype is a web application allowing mash-ups, collecting, editing, uploading, geovisualization, sharing using FOSS4G. The functionality of the developed system is: **a** allow users to collect geotagged multimedia data (e.g., text, location, image, audio) on the field; **b** store and manage collected geospatial data into spatial database server as a repository and publish it on the web through GIS web services; and **c** set up client interface for geovisualization (e.g., web, desktop). During this study, the participatory GIS is developed using ODK suite which offers a set of tools to manage mobile data collection. It is composed of three modules: ODK Build, ODK Collect and ODK Aggregate. On the basis of these modules, it is divided into the client (ODK Build, ODK Collect) and the server (ODK Aggregate) components. The ODK Build module is a graphical drag and drop interface for creating simple forms. Here, the field data collection for validation of LULC map is carried out where users need to provide basic information such as the unique_id, date, GPS location, LULC classes (i.e., forest, agriculture, settlement, water body, etc.) and image. The ODK Aggregate module is deployed using Apache Tomcat Server and the administrator allows creating new users and managing their privileges (e.g., export, view, upload

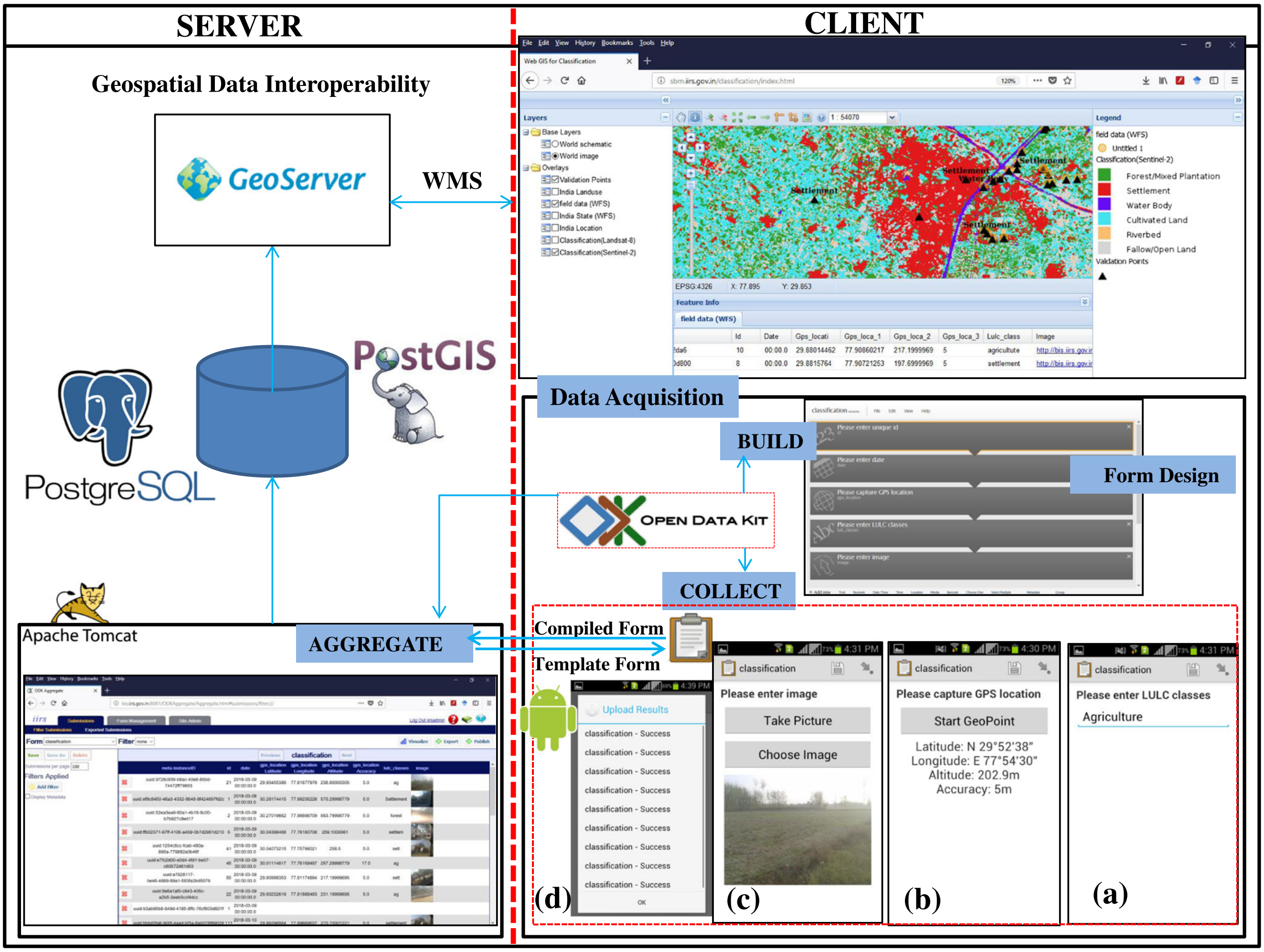


**Fig. 3** The system architecture of participatory GIS. The field data collection is carried out in the real-time environment for validation of the derived results. The screenshots of ODK collect App on the smartphone are reporting LULC information: **a** LULC class, **b** GPS location, **c** Captured image, **d** Upload results to the online repository of ODK Aggregate server. The geospatial data results are then published in GeoServer environment. The server is designed for interoperability and allows publishing geospatial data using OGC standard (e.g., WMS, WFS). The distributed GIS client can now access the network-enabled data for geovisualization and querying

and delete information). The ODK Collect module allows users to perform field data collection using mobile device. It makes data collection easy by including get blank form, fill blank form, edit saved form, send finalized form, view sent form, etc. The collected and compiled field information is stored to spatial database server (e.g., PostgreSQL and PostGIS). The geospatial attributes of spatial database server is published in GeoServer environment for interoperability using OGC standards. Here, OGC Style Layer Descriptor (SLD) standard is published in GeoServer to manage the appearance of geospatial data. The distributed GIS-based viewer is developed to geospatial data visualization and querying. The WMS layer represents LULC classes with name and symbol. When clicking on a point feature, a WFS GetFeatureInfo request is performed to retrieve collected information. The suitability of the participatory system is tested in LULC map validation, i.e., field data of LULC classes, location, images. The collected field data can be used for accuracy assessment of LULC map.

## Experimental results

Web-based geoprocessing standard (OGC WPS) provides a standardized service interface to publish and perform geoprocessing and analysis over the network. The SOA based scientific geoprocessing of unsupervised classification is carried out with the framework of 52° North WPS4R. The framework is available as open source implementation and supports multiple processing backends, i.e., WPS4R, GRASS GIS, Sextante Spatial Data Analysis Library, etc. WPS4R processing backend depends on annotated R-script of the algorithms. Initially, the R-script of unsupervised classification is annotated by setting inputs and outputs parameters. The algorithm adopts high-performance computing in order to process large datasets quickly and efficiently. The speedup and efficiency of the algorithm depend on the multi-core in the system. For example, the four cores (multi-core) of the system are considered for parallel processing. In order to develop a parallel computing framework, the remote sensing datasets are subdivided into four smaller tiles to be run in parallel and each tile is assigned to an independent processor, where the processing object computes the algorithm independently. After that, it merges the four independent derived results into a single and final result. The parallel computing with R is carried out with the libraries, i.e., parallel, foreach, doParallel, etc. In the experiment of unsupervised classification, the framework is processing only one algorithm while it is interoperable, reusable and scalable for different geoprocessing algorithms for remote sensing and GIS applications. The online geoprocessing experiment is performed for the unsupervised algorithm computation for the satellite images (Sentinel-2, Landsat-8) of Haridwar district of Uttarakhand state in India. However, the algorithm and framework are location independent for any type of geoprocessing. In order to compute algorithms, the important steps are described in three scenarios as discussed below:

In the initial scenario, the input datasets are published as OGC web service and integrated into the algorithm. The input datasets viz. polygon geometry of Haridwar district of Uttarakhand state in India; Sentinel-2 datasets are integrated as WFS and WCS, respectively in geoprocessing algorithm. Apart from this, it can also be uploaded as local system files in the server repository for geoprocessing. The pre-processing GIS operations, i.e., crop/mask, resampling, etc. are carried out to bring the datasets in similar resolutions. In the second scenario, the annotated R-script of unsupervised classification is generated with the input datasets (Sentinel-2, Landsat-8). The derived output of unsupervised classification using Sentinel-2 datasets is shown in Fig. 4.

The automated raster symbology operation is performed with the classification maps. Hence, the output is symbolized in the classes such as forest/mixed plantation, settlement, waterbody, cultivated land, riverbed and fallow/open land. Another annotated script using Landsat-8 dataset is computed for unsupervised classification. Similarly, the other output is also symbolized as similar classes that of Sentinel-2 data. Figure 5 represents the unsupervised classification result using Landsat-8 datasets. Once geoprocessing of the algorithm is complete, the output is stored in the specified working directory and then return to the user's browser for geovisualization and downloading. For geovisualization, the user can access the output map in an interactive and responsive environment. The online geovisualization of the outputs is demonstrated in Fig. 6. The detailed description of the classes and corresponding area with computation time is shown in Table 1.

It is observed that the computation time of unsupervised classification algorithm with Sentinel-2 datasets is more as compared to Landsat-8 dataset. It has occurred due to large number of pixels in Sentinel-2 data as well as resampling operation to obtain similar resolution. In addition, the algorithm computes different percent area of LULC classes in the same landscape. It has occurred due to variation in spectral characteristics into clusters of the various types of remote sensing datasets.

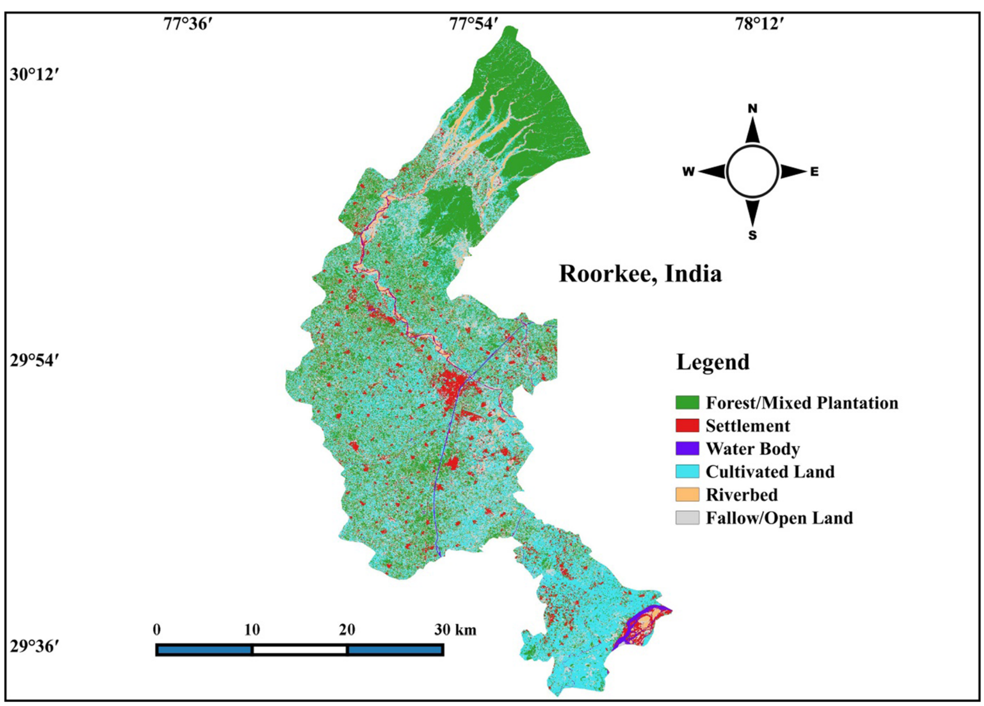


**Fig. 4** Unsupervised classification using Sentinel-2 datasets

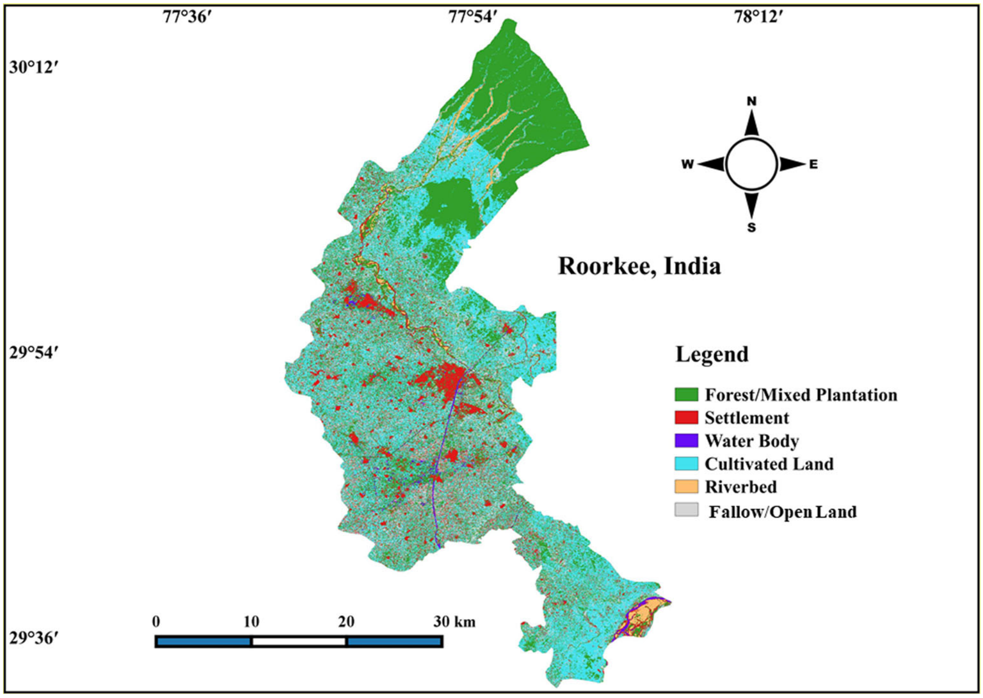


**Fig. 5** Unsupervised classification using Landsat-8 datasets

The most widespread classification accuracy technique is error matrix, which is used to describe a series of analytical and descriptive statistics (Smits et al. 1999). This method is very useful to represent accuracy assessment of classification map. The accuracy of a LULC classification map is the degree to which the map is acceptable with the reference LULC classification (e.g., ground truth). The users of classification maps need to know how accurate the derived maps are in order to use the geospatial map information more efficiently and correctly. Here, the accuracy assessment is performed by comparing the map generated by online geoprocessing and analysis to reference geospatial data collected using participatory GIS approach. The validation of derived classification map is carried out with the overall accuracy, users, accuracy, producers

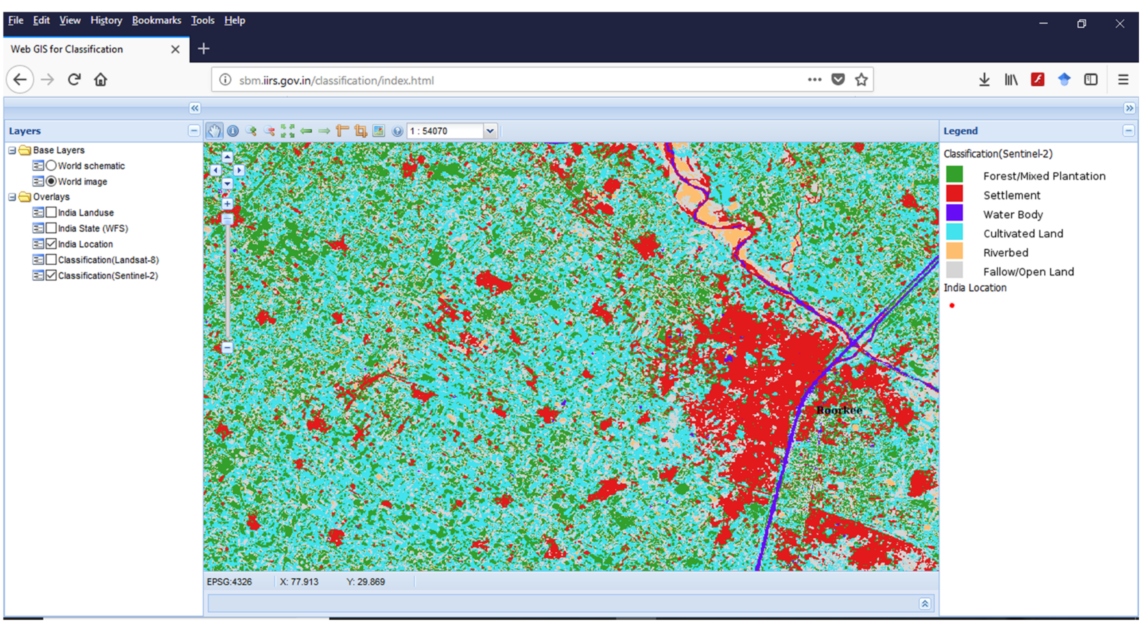


**Fig. 6** A distributed GIS-based framework of the derived output of unsupervised classification. It provides an interactive/responsive user interface for geovisualization, geocoding, etc. The experimental result is published as GIS web services with legend panel. It automatically computes raster symbology of the derived results. Here, the democratization of unsupervised classification (Sentinel-2) result is mapped with obtained LULC classes

**Table 1** The results of unsupervised classification system (area and computational time)

| Classes | Area% (Sentinel-2) | Area% (Landsat-8) | Landsat-8 | | Sentinel-2 | |
|---|---|---|---|---|---|---|
| | | | Sequential | Parallel | Sequential | Parallel |
| Forest/Mixed Plantation | 36.45 | 37.50 | 76.34 (Second) | 30.67 (Second) | 548.65 (Second) | 219.21 (Second) |
| Settlement | 7.01 | 10.24 | | | | |
| Waterbody | 0.90 | 0.96 | | | | |
| Cultivated Land | 38.06 | 36.04 | | | | |
| Riverbed | 4.25 | 1.43 | | | | |
| Fallow/Open Land | 13.38 | 13.80 | | | | |

accuracy and Kappa statistics. The accuracy assessment parameters of the derived classification maps are shown in Table 2.

## Discussion

In general, most web GIS applications focus on data dissemination, geovisualization and querying. The web GIS based geoprocessing system for remote sensing and earth observation data processing and analysis have been customized on a COTS platform and have the limitation of adaptability, platform dependency, etc. To address these problems, online image processing system using open source technology has been implemented and demonstrated during this study to provide an environment for real-time geospatial data processing, reusability, flexibility, scalability and distributed computing. The most suitable geoprocessing framework 52° North WPS4R comprises statistical computing using R programming and WPS for online geoprocessing. Based on the literature review, the proposed geoprocessing framework is updated and is suitable for the implementation and demonstration of remote sensing algorithms. The WPS4R based on OGC WPS standard and R backend provides an integrated environment for geoprocessing in a standardized way. The developed geoprocessing framework is generic and dynamic in nature and can be applied for other geoscientific applications. In addition, pre-and post-processing operations are automatically performed by the developed system, which is most time and resource consuming task in any modelling and analysis of geospatial data. The experimental run of the algorithm of unsupervised classification for study area has illustrated that the interoperable system has been able to acceptably quantify LULC classes, i.e., forest/mixed plantation, settlement, water body, cultivated land, riverbed and fallow land/open land. However, the algorithm requires some deep learning based optimization for extracting the LULC features in an efficient manner. As a result, the developed system has been successfully tested and evaluated with the latest version of browsers in the market such as Chrome, Firefox, IE, etc. Here, the online application is running with the hardware capability as Intel(R) Xeon(R) CPU E5–2630 v4 @ 2.20GHz (8 CPUs), ~2.2GHz, 16 GB RAM and Ubuntu 14.04 as OS. In order to improve the performance of the online application, parallel processing based algorithms of unsupervised classification are deployed under the environment of geoprocessing framework to cope with the large-scale remote sensing data requirements. In order to reduce redundancy in data and processing services, the datasets services (WMS, WFS and WCS) and processing services (geoprocessing of unsupervised classification) can be shared and consumed by other GIS clients.

Furthermore, participatory GIS is developed and integrated with distributed GIS environment for the validation of derived results of the unsupervised classification algorithm. The implemented architecture is developed using FOSS4G tools. The

**Table 2** Accuracy assessment of derived unsupervised classification maps

| Landsat-8 | | | Sentinel-2 | |
|---|---|---|---|---|
| LULC Classes | User Accuracy | Producer Accuracy | User Accuracy | Producer Accuracy |
| Forest/Mixed Plantation | 89.47 | 58.62 | 84.21 | 61.53 |
| Settlement | 57.89 | 68.75 | 63.15 | 80.00 |
| Water Body | 50.00 | 42.85 | 83.33 | 71.42 |
| Cultivated Land | 46.66 | 53.84 | 66.66 | 76.92 |
| Riverbed | 45.45 | 100.00 | 72.72 | 100 |
| Fallow/Open Land | 81.81 | 81.81 | 45.45 | 41.66 |
| Overall Accuracy | | 64.19 | Overall Accuracy | 69.13 |
| Kappa Coefficient | | 0.55 | Kappa Coefficient | 0.61 |

updated and real-time information of LULC classes (forest, agriculture, settlement, etc.) is accessed through participatory GIS system which is valuable for effective decision-making. The validation shows that the Kappa coefficient of classification result for Sentinel-2 is 0.61 and Landsat-8 is 0.55, which is acceptable in Kappa accuracy.

This online geoprocessing application demonstrates that the proposed approach is efficient to characterize LULC classes obtained from unsupervised classification of remote sensing data. The derived LULC classes demonstrate that if the input/output parameters and algorithm are appropriate, the derived outputs of online unsupervised classification system may have the same level of accuracy as the outputs of COTS software. The difference has generally occurred with the time spent on transferring and downloading geospatial datasets over the network. If the online geoprocessing system is accessed by multiple users simultaneously, the Internet connectivity is slow, or remote sensing datasets are at the large-scale, then the overall processing time will inevitably increase. With the experimental analysis, it has observed that the online image classification system has a good performance in a web browser environment with no interferences. However, the web applications may have unexpected circumstances such as performance issues, poor network connectivity, hackers attack, etc. This online application still requires the development of better GUI in an operational environment.

During the study, the cloud-free satellite imagery (Sentinel-2, Landsat-8) is downloaded and stored in the application server repository. The similar dataset is published as data service on the network and can be used for further geoprocessing and analysis. Here, the satellite data assessment and downloading process is not automated and is depending on human judgments and insight. Also, the unsupervised classification algorithm is extended for handling cloud mask and it can be considered for future implementation.

## Conclusion and future prospects

This paper presents a robust and interoperable geoprocessing framework with different modules for developing the distributed GIS-based unsupervised classification system. To demonstrate and validate the online geoprocessing, the unsupervised classification algorithm is implemented. The remote sensing images (Sentinel-2 and Landsat-8) are tested on the developed online system. The experimental and validation results identify that the developed system can accomplish the unsupervised classification of remote sensing datasets. The performance of web application is mainly influenced by network status and geoprocessing engine processing capabilities as it depends on the number of cores. The online system has achieved good computational speed through parallel processing based algorithm and it will be a valuable initiative for large-scale data processing. The online unsupervised classification system is helpful for common users with different conveniences. The user can access and compute the unsupervised classification of remote sensing data through updated browsers anytime and everywhere but Internet connectivity is required.

Due to SOA based geoprocessing framework, it has also become easy to share the geoprocessing services of unsupervised classification to the other GIS clients. It reduces the redundancy and allows the application developer to focus on GUI rather than processing backend. Hence, it exploits the power of distributed computing. This study discusses the advantages of online classification and demonstrates a model to turn the possibility into reality. The online classification model can be considered as a prototype of interoperable geoprocessing services. This geoprocessing service environment can also be integrated with other geoscientific applications such as natural resource management, agriculture yield estimation, etc.

In order to validate the classification results, an accuracy assessment of derived LULC map is performed using participatory GIS approach. The participatory GIS prototype is developed for providing access to field data collection, integration and accuracy assessment. It allows application users to use their mobile devices to collect geospatial data related to LULC references classes. The user-generated geospatial data is migrated into the spatial database and then published as a layer on the web for geovisualization and simple querying. In addition, more image processing algorithms can be deployed under the geoprocessing engine to serve the large group of users in the future. For further development, the algorithm of unsupervised classification can be optimized through deep learning methods. In order to improve the GUI of the classification system, more development is needed to handle the unexpected situation in an operational environment efficiently. Furthermore, this study proposed theoretical and practical aspects for geospatial community and developers to build operational and efficient geocomputation model in future.

**Acknowledgements** The first author would like to acknowledge MHRD, Govt. of India for providing the financial assistantship during the research and also want to extend his sincere thanks to IIRS/ISRO, Dehradun for their help in offering the resources in running the program. Thanks to Dr. S.K. Srivastav, dean (academics), IIRS/ISRO for their encouragement and technical support for this research study. Authors express their special thanks to Dr. A. Senthil Kumar, former Director IIRS/ISRO for sharing his overall suggestions to develop this web application.



## References

Alonso K, Espinoza-Molina D, Datcu M (2017) Multilayer architecture for heterogeneous geospatial data analytics: querying and understanding EO archives. IEEE J Sel Topics Appl Earth Observ Remote Sens 10:1–11

Astsatryan H, Hayrapetyan A, Narsisian W et al (2015) An interoperable web portal for parallel geoprocessing of satellite image vegetation indices. Earth Sci Inf 8:453–460

Bastin L, Buchanan G, Beresford A et al (2013) Open-source mapping and services for web-based land-cover validation. Eco Inform 14:9–16

Brovelli MA, Minghini M, Zamboni G (2008) Web based participatory GIS with data collection on the field - a prototype architecture. OSGEO Journal 13:29–114

Brovelli MA, Minghini M, Zamboni G (2015) Public participation GIS: a FOSS architecture enabling field-data collection. Int J Digital Earth 8:345–363

Castronova AM, Goodall JL, Elag MM (2013) Models as web services using the Open Geospatial Consortium (OGC) Web Processing Service (WPS) standard. Environ Model Softw 41:72–83

Chen Z, Chen N, Yang C, Di L (2012) Cloud computing enabled web processing service for earth observation data processing. IEEE J Sel Topics Appl Earth Observ Remote Sens 5:1637–1649

Dhamaniya A, Sonu M, Krishnanunni M et al (2016) Development of web based road accident data management system in GIS environment: a case study. J Indian Soc Remote Sens 44:789–796

Dubois G, Schulz M, Skoien J et al (2013) eHabitat, a multi-purpose web processing service for ecological modeling. Environ Model Softw 41:123–133

Flanders D, Hall-Beyer M, Pereverzoff J (2003) Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction. Can J Remote Sens 29:441–452

Giuliani G, Nativi S, Lehmann A, Ray N (2012) WPS mediation: an approach to process geospatial data on different computing backends. Comput Geosci 47:20–33

Haklay M, Singleton A, Parker C (2014) Web Mapping 2 . 0 : The Neogeography of the GeoWeb Web Mapping 2 . 0 : The Neogeography of the GeoWeb. 6:2011–2039

Hinz M, Nüst D, Proß B, Pebesma E (2013) Spatial statistics on the geospatial web. In: 16th AGILE conference on Geographic Information Science. Leuven,Belgium, pp 14–17

Karnatak HC, Shukla R, Sharma VK et al (2012) Spatial mashup technology and real time data integration in geo-web application using open source GIS - a case study for disaster management. Geocarto Int 27:499–514

Karnatak HC, Singh H, Garg RD (2017) Online spatial data analysis and algorithm development for geo-scientific applications using remote sensing data. Proc Natl Acad Sci India Sect. A Phys Sci 87:701–712

Lanig S, Zipf A (2009) Interoperable processing of digital elevation models in grid infrastructures. Earth Sci Inf 2:107–116

Li Z, Yang CP, Wu H et al (2011) An optimized framework for seamlessly integrating OGC web services to support geospatial sciences. Int J Geogr Inf Sci 25:595–613

Long W, Srihann S (2004) Land cover classification of SSC image: unsupervised and supervised classification using ERDAS imagine. In: IEEE International Geoscience and Remote Sensing Symposium, Anchorage, AK, USA, pp 2707–2712

Maisonneuve N, Stevens M, Ochab B (2010) Participatory noise pollution monitoring using mobile phones. Inf Polity 15:51–71

Maulik U, Sarkar A (2012) Efficient parallel algorithm for pixel classification in remote sensing imagery. GeoInformatica 16:391–407

Michaelis CD, Ames DP (2009) Evaluation and implementation of the OGC web processing service for use in client-side GIS. GeoInformatica 13:109–120

O'Reilly T (2007) What is web 2.0: design patterns and business models for the next generation of software. Commun Strateg 1:17

Porta J, Parapar J, García P et al (2013) Web-GIS tool for the management of rural land markets: application to the land Bank of Galicia (NWSpain). Earth Sci Inf 6:209–226

Qin CZ, Zhan LJ, Zhu AX, Zhou CH (2014) A strategy for raster-based geocomputation under different parallel computing platforms. Int J Geogr Inf Sci 28:2127–2144

Quirita VAA, Da Costa GAOP, Happ PN et al (2016) A new cloud computing architecture for the classification of remote sensing data. IEEE J Sel Topics Appl Earth Observ Remote Sens 10:409–416

R Development Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rathore DS, Chalisgaonkar D, Pandey RP et al (2010) A web GIS application for dams and drought in India. J Indian Soc Remote Sens 38:670–673

Rautenbach V, Coetzee S, Iwaniak A (2013) Orchestrating OGC web services to produce thematic maps in a spatial information infrastructure. Comput Environ Urban Syst 37:107–120

Singh H, Garg RD (2016) Web 3D GIS application for flood simulation and querying through open source technology. J Indian Soc Remote Sens 44:485–494

Singh H, Karnatak HC, Garg RD (2017) An automated and optimized approach for online spatial biodiversity model: a case study of OGC web processing service. Geocarto Int 6049:1–21

Smits PC, Dellepiane SG, Schowengerdt RA (1999) Quality assessment of image classification algorithms for land-cover mapping: a review and a proposal for a cost-based approach. Int J Remote Sens 20:1461–1486. https://doi.org/10.1080/014311699212560

Sun Z, Fang H, Di L et al (2016) Developing a web-based system for supervised classification of remote sensing images. GeoInformatica 20:629–649

Yue P, Guo X, Zhang M et al (2016) Linked data and SDI: the case on web geoprocessing workflows. ISPRS J Photogramm Remote Sens 114:245–257

Zehtabian A, Ghassemian H (2017) An adaptive framework for spectral-spatial classification based on a combination of pixel-based and object-based scenarios. Earth Sci Inf 10:357–368