POLITECNICO DI MILANO

School of Civil and Environmental Engineering

Polo Territoriale di Como
Master of Science in
Environmental and Geomatic Engineering

Thesis topic

Land Cover Validation game

Supervisor:
Prof. Maria Antonia Brovelli
Assistant Supervisors:
Irene Celino
Dr. Monia Molinari

Masters of Science thesis by: Vijay Charan Venkatachalam
Matriculation number: 813508
Academic Year: 2014/2015

Acknowledgement

This Land Cover Validation game consumed huge amount of work, research and dedication. Still, implementation would not have been possible if we did not have a support of many Citizen Scientist (Crowd from FOSS4G). Therefore, I would like to extend my sincere gratitude to all of them.

First of all I am thankful to CGR. S.p.a for providing free Orthophoto of the study area. It was a great help from them. I would like to express my sincere thanks to Prof. Maria Brovelli, Queen of FOSS4G for giving me this great opportunity to work on this research thesis and for guiding me throughout this thesis.

I would like to express my special thanks of gratitude to my assistant supervisor of this thesis work Irene Celino (Senior Specialist @ CEFRIEL) and Dr Monia Molinari (Post-doc in Politecnico Di Milano) who supported me in this research thesis. I also would like to convey special thanks to Andrea Fiano (Senior Developer @ CEFRIEL) and Simone Corti (Technical Staff @ Politecnico Di Milano) for provision of expertise, and technical support in the implementation. Without their superior knowledge and experience, the game wouldn't have become a great success.

"Friends... they cherish one another's hopes. They are kind to one another's dreams"- Henry David Thoreau. It is very difficult to thank my friends in words because they provide moral support all these days. It is also difficult to list my friends since this space is not sufficient. I love me them all. Last but not the least; I would like to thank my parents and relatives for the complete support.

Effort never fails

The best is yet to come

Abstract

Land Cover is often a critical variable in many researches, including climate modelling, biodiversity monitoring, environmental and sustainable development. As the result of various freely available land cover data, it has become easier to assess the environmental variables. But the problem is how good these data for the particular application are. Though there are many techniques for assessing the quality of the land cover dataset, some innovative solution is needed to improve the quality of such data. This research thesis deals with such innovative solution to validate the land cover data sets.

The different validation techniques help in improving the quality. The validation performed by comparing the higher quality land cover dataset with another land cover dataset is considered as the ideal one, such comparison has been focused in this current research thesis. The comparison between the regional Italian land cover dataset with global land cover dataset is performed for quality improvement, but this quality may not be sufficient for some of the land cover applications.

The main objective of this research thesis is to develop an interactive WebGIS with a gaming approach for the Crowdsourcing-based validation of GlobalLand30, which is the global land cover dataset at 30 meter. The disagreement pixels which are obtained from the comparison of regional Italian land cover with Global Land 30 is fed into the game for further validation by means of the Human-based computation technique (Game With A Purpose).

The validation is performed by the participants of FOSS4G Europe conference pixel by pixel. The instant results are obtained in the game which gives percentage of Italian land cover data agreement, percentage of the Globe Land 30 agreement, percentage of other categories, and contribution made by the players. This analysis has been fully completed for Como city area for the year 2000. This research thesis proves that the combination of open source software with Crowdsourcing gives rise to innovative application which solves the major environmental problems.

Keywords: Land cover validation, WebGIS, Game With A Purpose, Crowdsourcing.

Table of Content

List of Figure

List of Tables

List of Equations

ACRONYMS

Chapter 1

1. Introduction

Land cover data represents very valuable resource for many different studies related to the environment and sustainable development. However, the usage of this data in many applications cannot disregard its validation and the knowledge of its classification accuracy. Among the land cover validation techniques, Web-based applications are growing increasingly. The approaches proposed by Fritz et al. (2009)[i] and Bastin et al. (2013)[ii] represent notable examples. Geo-Wiki (Fritz et al., 2009)[i] is certainly one of the most popular applications used for validating global land cover using crowd sourcing. The disagreement maps between different global land cover products are provided, where users validate these products using Google Earth. In the application of Bastin et al. (2013)[ii], authorized users have to visually assess land cover and may provide uncertain information at various levels: from a general rating of their confidence to a quantification of the proportions of land-cover types within a reference area.

The objective of this work is to develop an interactive WebGIS for the Crowdsourcing-based validation of GlobeLand30, a new global land cover dataset at 30 meters resolution derived from the classification of Landsat (TM and ETM+) and HJ-1 satellites images according to the pixel-object-knowledge-based (POK) based approach (Chen et al., 2014)[iii]. The dataset has been produced by the Chinese government and released as open data in September 2014. It is available for the two baseline years of 2000 and 2010.

Within a research study aimed to evaluate the classification quality of GlobeLand30 on the Italian area (Brovelli et al 2015)[iv], this dataset has been compared with more accurate Italian land cover maps. Results show a degree of disagreement that ranges between 10% and 20%. For the non-coherent data another level of validation is needed.

The implemented WebGIS is intended to involve citizen scientists to classify the non-coherent pixels. In comparison to Geo-Wiki, which makes use of Google Earth in its Web application for the validation purpose, we implemented a Web application which makes use of high resolution aerial photos in the form of a gaming environment which may attract many more citizens for the validation process. Currently, the area of Como municipality (Lombardy, Italy) has been considered for this WebGIS implementation.

1.1 Global land cover assessment

A long recognized fact is that anyone can make a map of any place. The key question is how accurate is the map for that particular application? One perspective of this problem is that a map is nothing more than a hypothesis until it is tested against higher quality reference data (Olofsson et al., 2012)[v]. It is this perspective that drives the global scientific community to demand robust and continuous accuracy assessment of existing and future land-cover products.

1.2 Validation as a process

Validation may have several components. Most of the validation process has been described in details here.

Statistical Observation: The most important observation of accuracy is by means of statistical observations, which are unbiased estimates of accuracy measures and the variance of these estimates that are derived by probability sampling. These are probably the most useful general parameters, since they allow a user not only to weigh the magnitude of an error, but also the impact of error estimates or other process using the land cover information as an input. The way to estimate accuracy, using statistical observation is made possible using three parts, which includes the response design, the sampling design and the analysis lays.

Confidence maps: A classification algorithm will often provide a measure of confidence that quantifies how closely a classified observation matches the exemplars of the training set. This kind of confidence measure will tend to follow true accuracy if the training set is extensive and well-selected (McIver and Friedl, 2001)[vi].

Other comparisons: The other useful way to validate the land cover maps is to compare the target land cover map with other sources of land cover data. For example, a low confidence map can be compared with higher confidence level map. Here higher confidence level map acts as ground truth.

Qualitative-systematic accuracy reviews: The systematic review of the global land cover map is another useful approach for the assessment of accuracy. In this method the map is divided into regular sub regions, for example, on a latitude-longitude grid, and each sub region is examined separately to determine its accuracy. This examination is typically qualitative, using existing map sources, imagery, and expert knowledge to assess the map within the sub region.

Validation of land cover change: Validation of land cover change presents its own unique set of problems. It is easy to validate errors of commission by examining pixels that are identified as having changed, but because change is relatively rare, it is hard to validate errors of omission among large numbers of pixels that are identified as unchanged.

Citizen Science for land cover validation: Sparks., et al (2009)[vii] described Citizen Science Land Cover Classification based on Ground and Aerial Imagery. Results suggest that across methods in both ground-based and aerial based experiments, there are similar patterns of agreement and disagreement among participants across land cover classes. Bastin et al., (2013)[ii] and Fritz et al., (2009)[i] proposed a web based tools to validate the land cover maps.

1.3 Using Existing data in Global Accuracy Assessment

The existing data denotes the reference data available to the accuracy assessment that would not require expending resources for field visits, imagery, or other reference data materials. Some effort might be needed to convert the data for use in the accuracy assessment, for example to reclassify the data to match the map nomenclature. Typically the existing data will lack the necessary coverage to serve as sole source reference data for a global map accuracy assessment, so it is second use, supplementing the accuracy assessment sample that is the most likely application.

There are several considerations which are relevant to incorporating existing data into a global accuracy assessment protocol. The existing data must be first evaluated to determine whether they are compatible with the response design protocol, which includes classification scheme, spatial support of reference data which are specifically for the map. If this condition is satisfied, the next step would be to determine the sampling design if any, that was used to collect these data. The ideal situation is that the existing data originated from a probability sampling design. Ongoing environmental monitoring programs such as the National Resources Inventory (Nusser and Goebel, 1997)[viii] and Forest Inventory and Analysis (USFS, 1992) in the United States are potential sources of high quality reference data originating from a probability sampling design, but problems of data confidentiality and administrative coordination may still be considerable. The design-based framework can be achieved using a dual frame sampling-estimation method by combining two probability samples. If the existing data have not been collected using a probability sampling design, their use may not be representative of the larger population.

There is no additional field or data acquisition costs incurred when obtaining existing data, these data nevertheless still have costs associated with their use in a global accuracy assessment. This cost is attributable to the time expended to determine if the existing data are compatible with the reference data being collected and to develop and implement the more complex estimation procedures. Then, the time and effort required to administer the exchange of data may be considerable, and confidentiality concerns may limit access to the data. Incorporating existing data into a global accuracy assessment merits consideration, but these data are more likely to play a minor role providing limited-purpose supplemental information rather than serving as a panacea for the significant problem of cost of a global accuracy assessment (Strahler et al., 2006)[ix].

1.4 Background

The recent advancements in engineering and technology have created a vast opportunity for the citizen science to have a significant impact on scientific research. There are innumerable instances among the research fields through the identification of galaxies, through the discovery of protein structures and the validation of land cover classes which is the Geo-Wiki project. This Geo-wiki project is the most recent example of a crowdscourcing effort to assist in environmental monitoring. The Geo-wiki project identifies locations where global land cover dataset disagrees on a given land cover classification. It makes use of crowd sourced participants, provides them with aerial imagery, and asks them to make a classification choice for that location of disagreement. This data shows a lot of promises in validating land cover datasets, but like most sources of citizen science and crowd sourced data; it fails to assure reliability and consistency.

In order to ensure reliability and consistency, most attempts to gather data come from more authoritative sources. The data from such authoritative sources, including Land Use/Cover Area Frame Survey (LUCAS) attempts to capture land use/cover data using trained surveyors, to collect and create land cover data rather than relying on novice citizens. These land surveyors personally visit many locations for the collection of land transect, taking photos, and determining land use/cover at a given location. In these authorities' sources, efficiency is sacrificed for reliability and consistency.

The data from the Citizen science need to be able to guarantee a relatively high amount of reliability and consistency, along with being efficient. The Geo-Wiki project makes use of Citizen Science to classify the land cover using the Google earth and aerial photos (Perger et al., 2012[x]; See et al., 2013[xi]; Foody et al., 2013[xii]; Comber et al., 2013[xiii] and Comber et al., 2014[xiv]). Comber et al., (2013)[xiii] focus on the differences between expert Geo-Wiki participants and non-expert Geo-Wiki participants when classifying land cover given aerial imagery. It also reports averaged agreement rates between participants of 66%-76% agreement when classifying land cover, observing that generally experts have higher max-minimum agreement than non-experts. Comber et al. (2013)[xiii] concludes with a similar result of experts being different than non-experts, but still class for "…further investigation into formal structures to allow such differences to be modelled and reasoned with".

Though the aerial photos have been available for some time, means of entry to quality datasets of ground-based photos have recently emerged. The Geo-wiki campaigns offer insight on humans' land cover classification using aerial photos. The other projects have attempted to test the effectiveness of using ground-based photos for humans' land cover classification. Although no research has tested these ground-based photos on a large number of data source when attempting to classify land cover (Iwao et al., 2006[xv]; Foody et al., 2013[xvi]). With the success of the Geo-Wiki project in contributing to growth of land cover datasets, OpenStreetMap (OSM) has also succeeded in contribution of environmental information from citizen science. OSM is an open source dataset that is built from citizens volunteering and creating geographic information. Arsanjani et al. (2013)[xvii]analysed the OSM accuracy and concluded that OSM as well as general forms of crowd sourced geographic data, can be reliable and consistent sources for mapping land use.

Prestopnik et al. (2011)[xviii]in his paper described about social computation system, where human participants and computer technologies work in coordination to produce results. Likewise, many citizen Science projects, human beings and computer technologies, each have a role to play in generating usable scientific data. Computer technologies are used to support and augment these human abilities, making them more efficient, accurate, and reliable. This social computation tasks are sometimes mundane or repetitive, but may also be complex, requiring specialized training or knowledge. The significant question that arises which is applicable for citizen science also is participant motivation. The motivation factor has been dealt detail in the chapter 3.

1.5 Outline of the thesis

In the chapter Citizen Science, the details about the rise of Citizen Science from the 19th century have been explained along with the development in environmental research, monitoring and policy making. The chapter 2 also categories the Citizen science by elaborating its key value as well as its practice in the real world. The brief of Volunteered Geographical information has been provided by considering its evolution, enabling technologies and its value. The chapter 3 deals with the Gaming literature, how gaming is implemented for Citizen Science and its evaluation. The chapter 3 also illustrates the design and how people get motivated because of the game design. The Chapter 4 (Land Cover) outline the data and methodology used to extract the disagreement pixels which are fed into the game in the validation process.

The chapter 5 mainly describes the design and architecture of the Land Cover Validation game, Web application development and the methodology employed within the game to validate the pixels. In the chapter 6, mapping party results along with the various analyses performed using the obtained results have been explained. Supplementary material and information is provided in the appendix of this thesis. Appendix A comprises of pre-processing of orthophoto and tuning Geoserver and Appendix B outlines Web Application functionalities. The source code is presented in grey shaded areas.

Chapter 2

2. Citizen science

2.1 Introduction

Alan Irwin, a social scientist used the term “Citizen Science” in 1995 to describe expertise that exists among those who are traditionally seen as ignorant 'lay people' (Irwin 1995)[xix]. The term “Citizen Science” has been re-defined by Rick Bonney as research technique that enlists the help of members of public to gather scientific data (Bonney et al. 2009)[xx]. On the other hand Roy et al., (2012)[xxi] defines it as simply the involvement of volunteers in science. In recent days, citizen science claims to be the knowledge of local environments, and education gained through experience, as well as the submission of scientific data by many online volunteers.

Introduction of new technologies made volunteers to make use of mobile internet and Smartphone application to increase accessibility and remote participation. By this means, use of familiar tools to report sightings of rare species or record noise pollution or uploading pictures of the event has been made possible in order to collect the global sources rapidly via the internet. With rising of online crowd sourcing projects termed as 'Citizen Cyberscience', there has been a growing acknowledgment of role that citizen science can play in participatory democracy and active citizenship (Rowland, 2012)[xxii].

The nature and application of Citizen Scientists is to structure the environmental policy, which also focuses on the educational and social impact of citizen science. The main key questions of Citizen Science have been addressed are about the quality assurance, how it benefits environmental monitoring and policy making.

The best recognised Citizen Science projects are Galaxy zoo, which makes use of volunteers to participate in astronomy research by classifying images of galaxies online. Originally, the images came from the Sloan Digital Sky Survey, an astronomical survey covering a quarter of the sky and over 930,000 galaxies (SDSS, 2013). The publicity of this project has been made via BBC radio and the BBC website. Volunteers were very useful in making numerous discoveries such as the first planet with four stars. The Swedish Bird Survey, which made use of Volunteer-collected data (i.e.) monitoring by a citizen scientists could prove useful in future assessments of wild bird populations and help to inform more targeted and efficient conservation efforts (Snäll et al., 2011)[xxiii].

In some of the projects related to environmental issues, the citizens may not have been previously aware of, or interested in. Such environmental activities are also labelled as citizen science often mould of 'public participation in scientific research' at lower levels of participation rather than alternative forms of knowledge. There are also many examples of tackling environmental issues by harnessing lay, local and traditional knowledge, which may not be accepted as citizen science projects and are therefore harder to identify.

2.2 The rise of Citizen Science

2.2.1 A short historical detail

Though the rise of Citizen Science took place in the 19th century, those projects were not recognised at that time. Prior to the 20th century, it was common for the work that we now refer to as science to be carried out by amateurs (Haklay, 2012[xxiv]; Rosner, 2013[xxv]). Charles Darwin (1809-1882), for example, had no formal training in science and yet is widely regarded as one of the most important scientists in history. Some concise historical context is needed in understanding the development of modern day Citizen Science and what is absolutely new about the current trend in citizen science. Modern day citizen science serve a return to a centuries-old approach to doing science, and to challenge the notion that science must be done by 'experts'.

The term 'Citizen Science' was coined by Alan Irwin in his 1995 book Citizen Science. Irwin describes how people accumulate knowledge in order to learn about and respond to environmental threats. Bonney., et al (2009)[xx] used the term 'citizen science' to refer to public participation in scientific research (Rosner., 2013)[xxv]. Irwin's definition is more focused on the form of knowledge beyond the scope of professional science, whereas second interpretation made by Rick Bonney has been focused to simple crowd sourcing.

"Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call."

Howe (2010)[xxvi]

The emergence of new technology such as mobile internet and Smartphone applications, to increase accessibility and remote participation, increased quickly more citizen science projects in the recent years. To be more practical, volunteers can make use of familiar tools to report sightings of rare species or record noise pollution, and that data from enormous global sources can be collected centrally and rapidly via internet. It is also necessary to know what today's citizen scientist can do to improve the way that environmental research is carried out and to better represent any threat to the quality of environmental science.

2.2.2 Development in environmental research, monitoring and policy making

In the few areas of Environmental Science, Citizen Science programmes give the pathway to the existing interests, dedication and in some cases, expertise, amateur enthusiasts and skilled professionals with no formal scientific qualification. Most of these programmes were not initially recognized as examples of citizen science.

One of the longest-running citizen science projects, the Audubon society's Christmas Bird Count began in North America with 27 dedicated 'birders' in 1900, which is available to anyone living in the designated survey areas, and with no specialist knowledge or experience is needed. Some of the programmes such as The Birds Directive where wild birds are protected and in many European countries these birds are monitored by chain of volunteers through the pan-European common Bird Monitoring Scheme, jointly led by Birdlife International and the European Bird Census Council (EEA, 2013).

Many of the projects are trying to focus on the local or large-scale environmental issues in which the citizens who get involved may not be aware of those issues or may not be interested. One particular Citizen Science project is "The Barneget Bay survey" in which biology teacher involved her students in the research which was focused on the degraded coastal habitat in their local area (Moore, 2011)[xxvii]. Such environmental activities doubtlessly marked as citizen science often considered as 'Public Participation in scientific research' which falls in Crowd-sourcing which is lower levels of participation (see section 2.2.3).

The Aarhus Convention[1] was framed by the United Nations Economic Commission for Europe (UNECE) during 1998 which gives rights to European citizens to take part in environmental decision-making. Recently, European Commission and European Environment Agency acknowledged the value of citizen science and lay knowledge for environmental research, monitoring and policymaking with the creation of the European Citizen Science Association (ECSA). The main goal of this organisation is to advance the knowledge about sustainable development and to engage the people of disadvantage group to

[1] *http://ec.europa.eu/environment/aarhus/*

take an active role in sustainable development to conserve and improve health and the environment.

2.2.3 Categorising citizen science projects

It's difficult to categories the citizen project due to a wide variety of potential subject areas, aims and approaches. However, comparing and contrasting the various citizen projects has been made possible by the author to figure out a project's success. Haklay's (2012) [xxiv] scheme classifies citizen science projects within a four-level framework of participation, based on the depth of their engagement with volunteers (Figure 1).

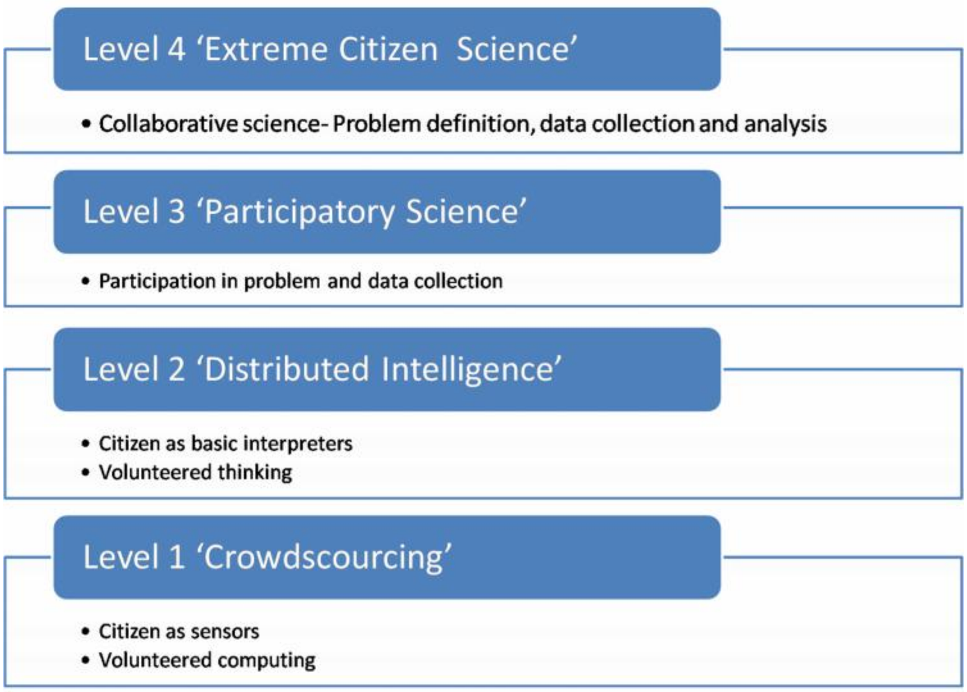


Figure 1: Participatory levels of Citizen Science. Source: Haklay (2012)[xxiv].

The level 1 is termed 'Crowdsourcing' which make use of volunteers simply as a means to collect data from distributed sensors or to provide computing power. The level 2 projects may provide participants with some basic skills before involving them in the collection and interpretation of data. The well-known examples of this level are Galaxy Zoo and eBird. The level 3 participants are more associated with the research oriented problem, definition and data collection. The level 4 projects are referred to as 'Extreme Citizen Science'. These projects involve the citizens who are the driving force behind the research and professional scientists are not involved at all. The level 4 is based on the level of engagement instead of encouraging judgements about how good specific projects are. The table 1 shows how Wiggins and Crowston classification, Roy et al classification, Haklay classification schemes can be used classify the different citizen science project.

Project	Brief description of project	Wiggins and Crowston classification	Roy et al classification	Haklay classification
Galaxy Zoo	Classifying images of galaxies	Virtual*	Mass* Contributory*	2 – distributed intelligence*
eBird	Collecting bird observations	Investigation*	Mass* Contributory*	2 – distributed intelligence
What's Invasive	Locating invasive plants	Conservation*	Mass Contributory	2 – distributed intelligence
ReClam the Bay	Restoring local bay's clams and oysters	Action*	Local Community-led	3 – participatory science
Corfe Mullen Bio-blitz	Identifying species in Corfe Mullen village and local area	Investigation / Education	Local Co-created*	3 – participatory science
Climateprediction.net	Volunteers' computers used to run climate prediction models'	Virtual*	Mass Contributory	1 - crowd sourcing

*Table 1: Classifying citizen science projects. *Indicates example classifications proposed by authors of classification schemes.*

2.2.4 Evolution of new technologies

In most of the Environmental projects which are termed as 'citizen science' involves citizen for data gathering and less utilized for setting aims and objectives. The growth of internet increased more online citizen science projects. One can make use of internet, Smartphone sensors and online or phone based games to influence how science and policy making are carried out (Graham et al., 2011[xxviii]; Haklay, 2012[xxiv]).

Many of the EU projects harness citizen as sensors by utilizing mobile technologies. In What's invasive, volunteers make use of an app that allows them to address sightings of invasive plants and animals by selecting species from the list which is shown and uploading the photos taken using their phones (Graham et al., 2011)[xxviii]. Using this app, 6000 observations of invasive species has been collected by 1900 registered users.

Online societies promote the inspiration for participating and continuing to participate in online citizen cyberscience. Citizen Cyberscience society is characteristic of a ranked order, unlike social networking sites, such as Facebook and Twitter, which are self-organising. (Wiggins and Crowston, 2011)[xxix], François Grey claims that the social interactions between some citizen cyberscience societies are identical to those of social networks and may provide a strong incentive for users to take part in it (Grey, 2009)[xxx].

2.3 The value of citizen science

2.3.1 Scientific Value

The value of citizen science is divided into scientific, education, social and policy aspects. The scientific value of citizen science is reliant on the quality of data collected and how these data are used. But for some citizen science projects, scientific data quality may not be the priority. Educational benefits and awareness-raising are also common aims.

2.3.2 Educational Value

The Educational value of citizen science reliant on the basic level of knowledge of those taking part admit that adult participants are often drawn to projects because of their existing interest as well as the level at which they are engaged in scientific content (Haklay, 2012) [xxiv].

2.3.3 Societal Value

Gollan et al., (2012)[xxxi] propose citizen science has the potential to bring the society near to science and to nature, bringing about a sense of ownership and help to create so called society which labours to guard its natural environment. By involving the volunteers to collect scientific data may greatly reduce the cost of scientific studies probably contributing better value of public money (Gardiner et al., 2012)[xxxii]. However, because projects vary widely in their goals and scope, it is difficult to generalise they may incur costs that do not apply to many traditional science projects, such as attracting contributors

2.3.4 Value of policymaking

Citizen science can actively involve in policymaking by providing opportunities to address environmental issues that directly affect citizens. Citizen science has been very useful to empower communities by involving them in research which can be used to drive forward policy changes (Rowland, 2012)[xxxiii].

2.4 Citizen science in practice

2.4.1 Motivation

There are many reasons for participating in citizen science projects and it may vary from project to project and from person to person. The key to the success of a project is to understand the motivations of contributors and project partners, but there can be a tendency to misunderstand or make assumptions about citizen scientists and their reasons for contributing (Grove-White et al., 2007)[xxxiv].

2.4.2 Framework

Many researchers have implemented frameworks for the design and management of citizen science projects for environmental studies which is summarised simply in the nine basic proposals for actions to be taken which was given by of Bonney et al.'s (2009)[xx] model (Figure 2), which was based on two decades experiences at the Cornell Ornithology Lab. Tweddle et al. (2012)[xxxvi] gave roughly similar and self-explanatory model which report on citizen science and environmental monitoring. Another framework offers a more considerate perspective on inputs and outcomes (Shirk et al., 2012)[xxxv] which was developed by the Conell Lab of Ornithology. Inputs are branched into scientific and public interests, which define the goals of the project. For example, some scientists may be keen to collect data about bird populations; some may be keener in cultivating the knowledge about bird biology, behaviour and habitats to general public. The first task is to identify and define the research questions from two different perspectives: first is to know whether citizen science project is best method to answer the research question which is being addressed (Tweddle et al., 2012)[xxxvi]; secondly to know whether the proposed research question is appropriate for citizen scientists (Bonney et al., 2009)[xx]. By either of this scientific questions, there is a need to consider what kind of scientific questions and which questions specifically, can be answered through citizen science (Haklay, 2012)[xxiv].

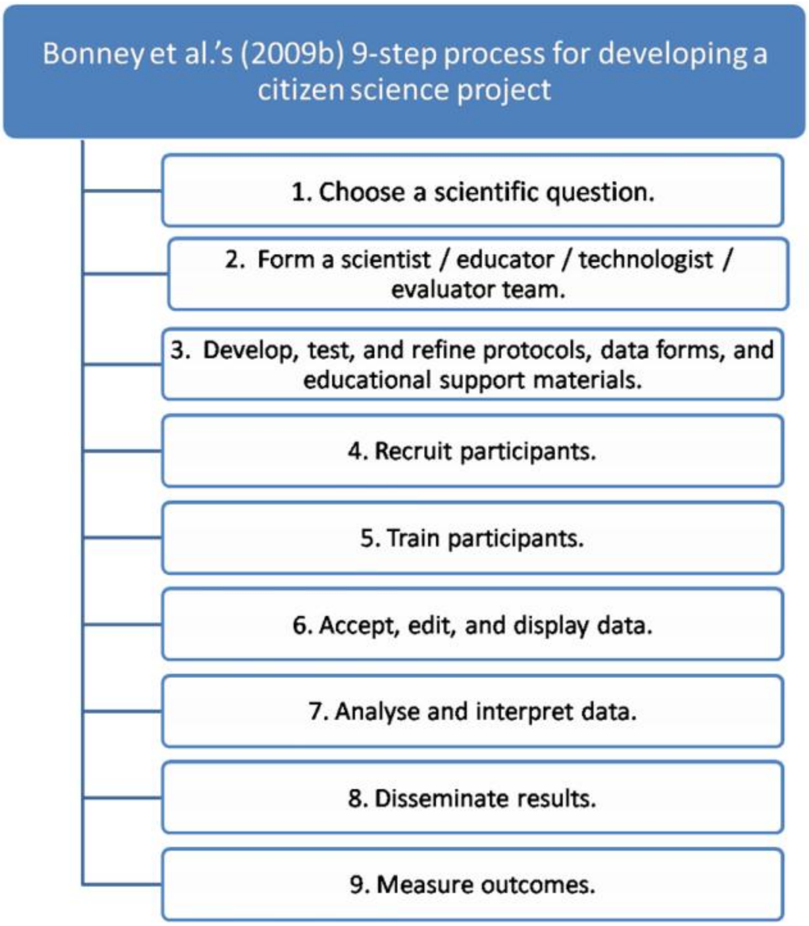


Figure 2: 9-Step process for developing a citizen science project (Bonney et al.'s 2009) [xx].

2.4.3 Guidelines

The process of setting up and managing a citizen science project has been given by Tweddle et al. (2012)[xxxvi], a practical guide freely available[2]. Mathieson (2013)[xxxvii] suggests teachers to recognize using citizen science in the classroom, highlighting those students should be contributing to real science, for locally relevant issues such as air quality or soil health. These frameworks and guidelines help to engage the public participation in scientific research model of citizen science, especially at lower levels of engagement, but makes clear that most 'extreme' and participatory form of citizen science is complex.

2.4.4 Quality assurance of citizen data

Based on the quality assurance methods employed, Citizen Science projects can be divided into two types: verified citizen science, in which collected information are verified by the experts; and direct citizen science, in which collected information are submitted without

[2] www.ceh.ac.uk/products/publications/documents/citizenscienceguide.pdf

verification (Gardiner et al., 2012)[xxxii]. These two approaches has been compared for a monitoring projects, where researchers found that with verified approaches, accurate estimates has been obtained using fewer volunteers. While direct citizen science was the cheapest method, verified citizen science has been proven to be more cost-effective in terms of data gathered per dollar than a traditional approach involving no volunteers.

Roy et al.’s (2012)[xxi] proposes that most of the environmental citizen science projects use quality assurance methods at some levels. The simple approach involves filters to remove data that differs from study's range due to time or geographical limits. More accurate approaches include testing and observation of study participants to organize common errors and how often those are likely to occur. In some citizen science projects volunteers were asked to take online tests during the project design stage.

In most of the citizen science projects the participants were given training before making them to participate in the project. They were also supported with teaching texts and field guides. In case of online citizen science projects, interactive tutorials, forums to discuss about the difficulties, frequently asked questions (FAQs) and videos are provided. Online training tools and video sharing websites might be very useful in supporting coaching for field studies, in order to revisit content and reach remote participants. In researchers point of view, quality assurance plays essential role in all citizen science projects and they believe it’s a way to increase confidence in the validity of scientific finding from citizen science projects (Delaney et al., 2008[xxxviii]; Golan et al., 2012[xxxi]).

2.4.5 Communication and recruitment

Some of the citizen science projects came up with the procedures for recruiting participants based on the different scenarios. Some large, funded citizen science projects benefit from the efforts of dedicated staffs. However, smaller and valuable projects may depend on scientists, teachers and community organisers with limited time and resources for recruitment and communications. In these cases, stakeholders can sought support and commitment while some reasonable uncertainties are accepted. The Cornell lab of Ornithology's citizen science recommend ideas on recruiting volunteers (Cornell University, 2013), which highlights how incentives and motivation is very useful in recruiting volunteers. This toolkit elaborates ideas of recruitment strategies which influence the participant profiles.

2.5 The World of Volunteered Geographical Information (VGI)

2.5.1 Introduction

Geographical information plays a vital role in many environmental applications. With internet connection one can provide information on a particular area on the earth's surface including linking other external sources. These things are made possible through Wikimapia application, which is significant in the creation of the Wikipedia encyclopaedia. Anyone with the Internet can view this information and can review it.

2.5.2 Evolving world of VGI

The most popular example of VGI is Wikimapia in which anyone can edit entries and the volunteers verify the results by checking the accuracy and significance. Different entries, including the description, pictures, and hyperlinks have been provided. Similar example is Flickr site that allows the users to upload photographs and locate them on the Earth's surface by latitude and longitude.

OpenStreetMap is another international effort which focuses on the creation of more elaborated representation of Earth Surface to create map data through volunteer effort. In the earlier days, it takes much time to create a simple map, but at present using sophisticated GIS tasks, it is possible for a child to create a simple map in 10 minutes. The ability to superimpose geographic information from different sources obtained by the Web. These things took away the old tradition of mapping by giving rise to more sophisticated technologies.

2.5.3 Enabling technologies

The technologies such as Web 2.0, Geotags, GPS, Graphics and Broadband communication made VGI easier and simple. The client can request for a particular task on the map, the instruction is sent to the server and the query is processed and sent back to the client (user). These processes are made simple using Web 2.0. Geotags have been inserted into many Wikipedia entries which are a standardized code which can be introduced to the appropriate geographic location. Of course the most important one is the Internet connection in order to access the information from World Wide Web (WWW).

2.5.4 Value of VGI

VGI acts as a very useful source of commercial intelligence and military. It helps in identifying the activities about life that takes place in nearby location which may be unnoticed by the world media. It also gives very interesting and fascinating value to the geographers.

Chapter3

3. Gaming for (citizen) science

3.1 Introduction

The people all over the world spend billions of hours in playing computer games each year. By doing this they spend a lot of time, money and consumes lots of energy. What if the money, time spent and energy consumed could be utilized for good purpose? (Ahn 2006)[xxxix]. What if the people, while playing computer games involve themselves to solve the real world problems without knowing that they are involved in solving real world problems?

Over several decades, we rely on computers to solve basic problems as well as massive computational problems. We started exploiting computers for everyday work. But still computers don't possess the basic conceptual intelligence or perceptual capabilities that most humans take for granted. What if we utilize human brains as processors in a distributed system? (Ahn 2006)[xxxix]. Yes, it's quite sure that each can accomplish a small part of a massive computation. This enormous potential can be achieved by "human computation" by tackling the problems which computer cannot achieve.

Scientific endeavours require data of very high quality. Prestopnik & Crowston (2009)[xl] demonstrated that Citizen Science efforts often produce high quality data, but the open question remains, how do the motivational techniques employed by Citizen Science projects attract the participants to get involved and stay involved. For sure Computer processors don't need any motivation to perform any work, but human brains need some sort of encouragement or inducement to do the work (Ahn 2006)[xxxix]. Games are the luring method for attracting people to participate in a task. Such games constitute a general mechanism for using brain power to solve an open problem. Designing such games is not quite easy. It is similar to designing the algorithm and it must be proven to be more efficient and correct. These "algorithms" which are similar to the silicon chip in the processor should provide human interaction with computers over the Internet.

The ESP games, the Google Image Labeller are the examples of Game with A Purpose (GWAP) in which the participants provide significant, factual labels to the images on the Internet as a post-effect of playing the game. These games are fast-paced, fun-filled and competitive which contributed more than 50 million labels. These labels were effectively used to progress Web-based image search which cannot be performed well with the computer-vision[3] techniques. This is an entertaining way to encourage players to label the images to improve the quality of web-based image searching. There are many more GWAPs such as pike-boom, which locates objects within the image; Phetch, which annotates images, with descriptive paragraphs; and Verbosity, which collects common sense facts in order to

[3] *Computer vision is a field that includes methods for acquiring, processing, analyzing, and understanding images and, in general, high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions.*

train reasoning algorithm. These games are more attracted to the people so that they help in solving scientific problems without getting bored of what they are doing.

3.2 Gamification

"The integration of the mechanics that make games, fun and absorbing into non-game platforms and experiences in order to improve engagement and participation" - Taylor Nelson Sofres (TNS) global[4]. Gamification term has been originated in the digital media industry, though it was documented in the early 2008, it filed extensive adoption in the late 2010, when many company players and conferences popularized it (Deterding et al. 2011)[xli]. The figure 3 describes about GWAP analyzed by the Gartner and M2 researches.

Figure 3: Gartner and M2 research analysis on Game With A Purpose.

[4] *TNS (formerly known as Taylor Nelson Sofres) is a leading market research and market information group. Formerly listed on the London Stock Exchange and a constituent of the FTSE 250 Index, the firm was acquired by WPP Group in October 2008 for 1.6 billion pounds.*

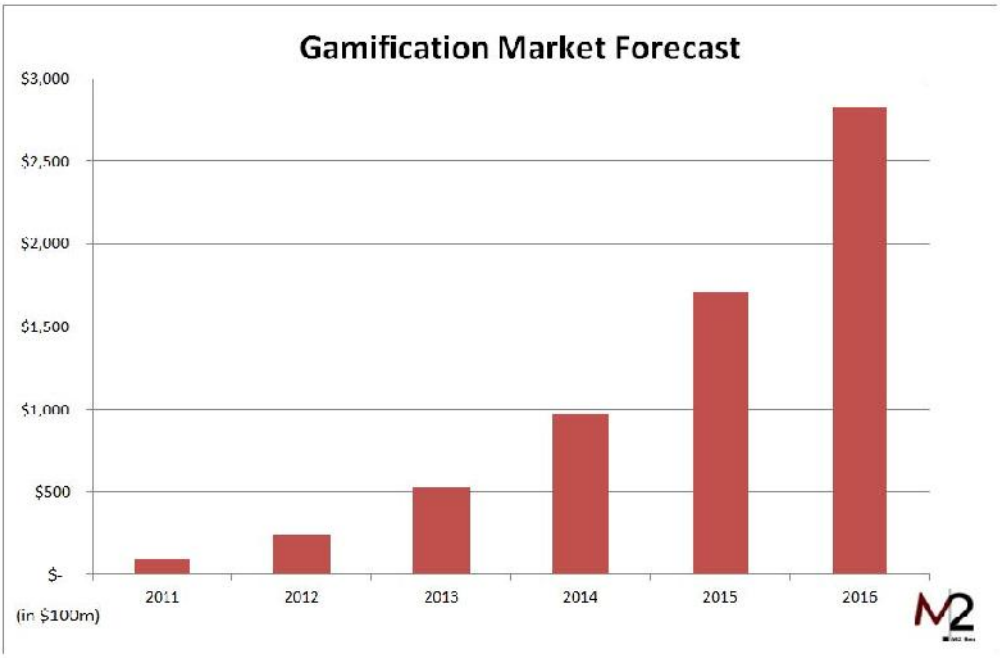


Figure 4: Gamification Market Forecast.

The figure 4 gives the forecast of how the Gamification projects grow in the near future .The classic definition in game studies that games are characterized by rules, and competition or strife towards specified, discrete outcomes or goals by human participants. Deterring et al. (2011)[xli] describes Gamification as the use of game design elements in non-game contexts (Figure 5). While games are usually played, play represents a different and broader category than games. They also distinguish gamification from playful interactions, playful design, or design for playfulness; they also assume that the design of gamified application will often give rise to playful behaviours and mindsets.

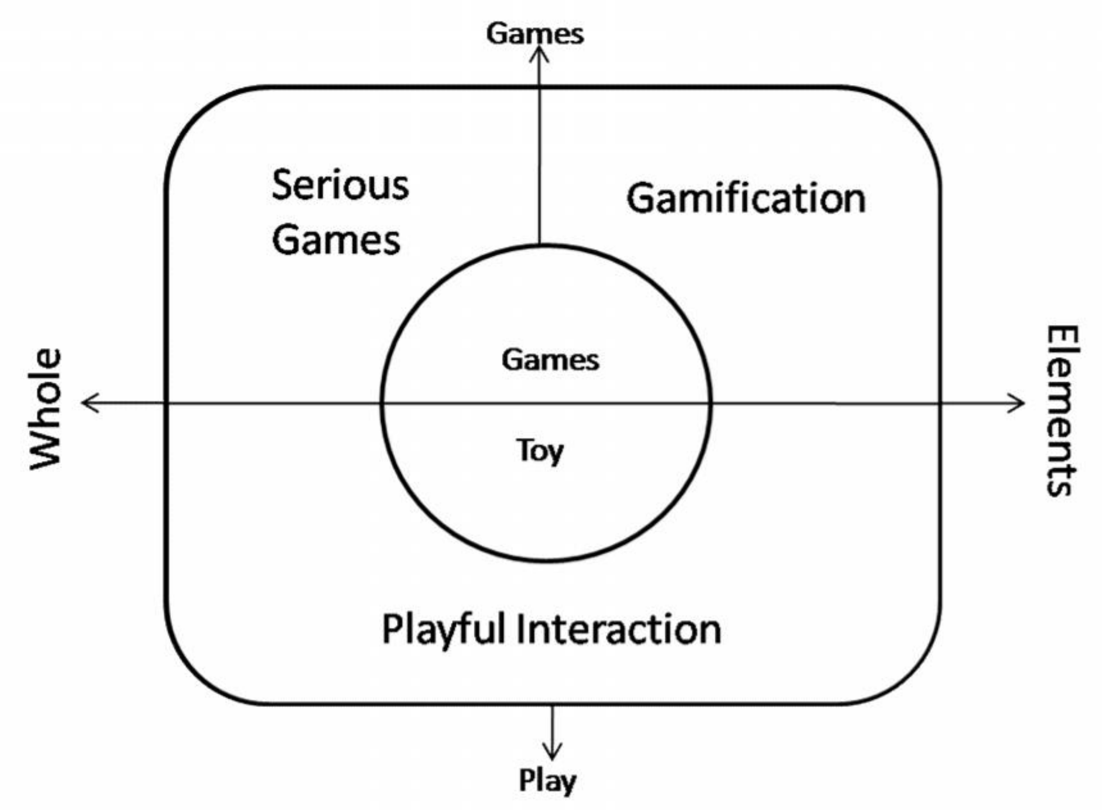


Figure 5: Gamification design.

There have been hardly any academic attempts at a definition of gamification. Current uses of the word seem to fluctuate between two major ideas (Deterring et al. 2011) [xli]. The first is the increasing societal adoption and institutionalization of video games and the influence games and game elements have in shaping our everyday life and interaction. The second, more specific idea is that-since video games are explicitly designed for entertainment rather than utility they can demonstrably produce states of desirable experience, and motivate users to remain engaged in an activity with un paralleled intensity and duration. Thus, game design is a valuable approach for non-game products, services, or applications which are more enjoyable, motivating, and/or engaging to use. The (Table 2)[5] describes categorization enterprise gamification in terms of infinite, finite, in system, out system, non-game context and game context. The (Table 3)[6] describes categorization and comparison of gamification with other gameful approaches by looking at characteristics such as spontaneity, rules, or goals.

[5] *https://en.wikipedia.org/wiki/Gamification*

[6] *https://en.wikipedia.org/wiki/Gamification*

Enterprise Gamification (Koivisto and Hamari 2014)					
	Gamification	Simulation	Serious game	Advergame	
Examples	SAP Community Network Stack Overflow Yahoo! Answers LinkedIn Amazon.com MySugr Duolingo Zombies, Run!	ERPSim	Merchants Triskelion Ribbon Hero		In System
Examples	Gamification Guru Leaderboard	Farm Simulator Trainz Surgeon Simulator 2013 Emergency Simulator	SAP Roadwarrior Stroke Hero Ten Euro Tetris The Accounted	Magnum Pleasure Hunt Coke Zero / James Bond	Out System
	Infinite	Finite			
	Non-game context		Game context		

Table 2: Categorization Enterprise Gamification.

	Play	Game	Serious game	Simulation	Gamification	Enterprise Gamification
Spontaneous	Yes	No	No	No	No	No
Rules	No	Yes	Yes	Yes	Yes	Yes
Goals	No	Yes	Yes	Yes	Yes	Yes
Structured	No	Yes	Yes	Yes	Yes	Yes
Real World Outcome	No	No	Yes/No	Yes/No	Yes	Yes
In System	No	No	No	Yes/No	Yes	Yes

Table 3: Categorization compares gamification with other gameful approaches by looking at characteristics such as spontaneity, rules, or goals.

3.3 Human Computation and Game With A Purpose (GWAP)

The emerging research area that centres on harnessing human intelligence to solve computational problems which are beyond the scope of existing Artificial intelligence algorithms is human computation. Human computation can influence the abilities of an unprecedented number of people via the Web to perform complex computations. The figure 6 shows the human computation cluster Venn diagram which describes how the human computation is developed from the other components.

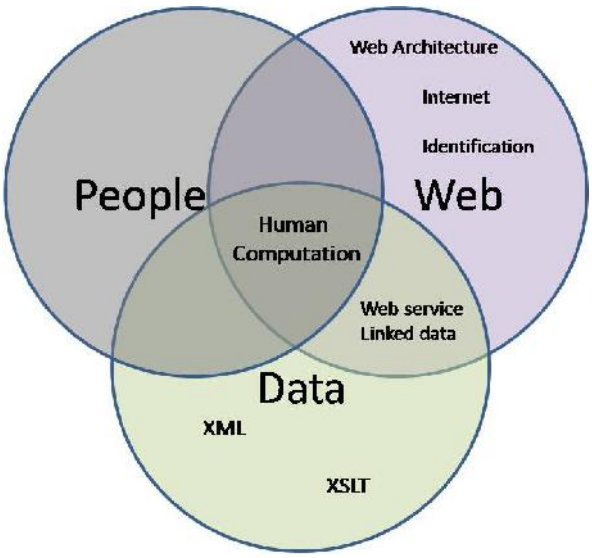


Figure 6: Human Computation Venn diagram.

There exist many real world problems of uncertainties which cannot be addressed or solved even using more sophisticated computer algorithms and for certain problems, no Artificial intelligence (AI) algorithm can exceed human performance, which includes perceptual tasks (object recognition, music classification) natural language analysis (sentiment analysis, language translation) and cognitive tasks (planning and reasoning). We will see how Human Computation helps in solving such problems.

There exist many human computation approach projects, but *ReCAPTCHA* and *ESP game* are very notable one because of their importance. *ReCAPTCHA* is human computation approach. The books written before the computer age are currently being digitized using Optical character recognition (OCR)[7], 30% of words cannot be recognized by OCR. Such unrecognized texts were added to *reCAPTCHA*. In order to verify the user answer, *reCAPTCHA* displays two words to the user. One word is a control word for which answer is known. If the user correctly types the control word the system assumes that the user is human and gains confidence that the unknown word is typed correctly. The experiment with 50 scanned NY Times Articles shows standard OCR accuracy of 83.5% and *reCAPTCHA* accuracy 99.1%. *ESP game* is another example of human computation approach which helps in labeling the images which are available on the web. Here each pair of players will be displayed with an image for which both players have to agree with the same answer.

3.4 Design of the Game

3.4.1 Introduction

These games are designed in a specific way to entertain people as well as to maintain the quality of the data collected. Ahn and Dabbish (2008)[xlii] articulate three GWAP games "templates" representing three general classes of games containing all the GWAPs they created up to date. Each of these templates defines the set of basic rules and winning condition that can be applied to any computational problem to construct a game that encourages players to solve problem instances. Finally, they also proposed a set of metrics defining GWAP success in terms of maximizing the utility obtained per human-hour spent playing the game. Sebastian Deterring et al. (2011)[xli] consider the term gamification for references to design elements, not game-based technologies or practices of the wider game ecology and they also suggest including the five levels which are given below:

Interface design patterns such as badges, levels, or leader boards (Sawyer and Smith 2008)[xliii]

Game design patterns (Bjork et al. 2005)[xliv]or game mechanics (McGonigal 2011)[xlv]

[7] Optical character recognition (OCR) is the mechanical or electronic conversion of images of typed, handwritten or printed text into machine-encoded text. It is widely used as a form of data entry from printed paper data records, whether passport documents, invoices, bank statements, computerized receipts, business cards, mail, printouts of static-data, or any suitable documentation.

Design principle or heuristics: guidelines for approaching a design problem or evaluating a design solution

Conceptual models of game design units, such as the MDA[8] framework (Hunicke et al 2004)[xlvi], Malone's challenge, fantasy, and curiosity (Malone 1981)[xlvii], or the game design atoms described in Braithwaite and Schreiber.

Game design methods, including game design-specific practices such as play testing and design processes like play centric design or value conscious game design (Belman and Flanagan 2010)[xlviii].These five levels are very necessary to design GWAP.

3.4.2 Desire to be entertained

Though there are many other projects that try to make use of individuals to perform some task by providing financial incentives, they don't entertain the individual. The current research work describes that GWAP approaches rely on the human desire to be entertained. GWAP should be correlated with a computational problem and therefore generates an input-output behaviour.

A GWAP's rules should encourage the players to perform the task properly in order to solve the computational problem. The problems which are defined should be easier for the humans, but difficult or impossible for the computers. The initial step is converting the problem into a GWAP which is the game with its structure including the rules and winning condition. The rules should also briefly explains the functionalities which can be done and can’t during the game.

3.4.3 Increase Player Environment

As described by Ahn and Dabbish (2008)[xlii], the most important aspect of GWAP is that the output is obtained in a way that is designed to be enjoyable. This should create an entertaining environment to the Player. The most of the challenges to increase the player’s motivation are translated into game features like timed response, score keeping (which includes leader board and best last player), player skill level. The features to enhance the player’s motivation are listed and explained as follows:

Timed response: Introducing the timed response increases the challenges into the game. The players are asked to choose the categories for the pixel in each round. This timed response establishes an explicit goal. Ahn and Dabbish (2008)[xlii] believes that they know from the literature on motivation in psychology and organizational behaviour that goals that are both well-specified and challenging lead to higher levels of effort and task performance than goals that are too easy or vague. It is vital to know the number of tasks for

[8] *In game design the Mechanics-Dynamics-Aesthetics (MDA) framework is a tool used to analyze games. It formalizes the consumption of games by breaking them down into three components - Mechanics, Dynamics and Aesthetics.*

players to complete within a given time period. The time period is calibrated to introduce challenge, therefore the time limits and time remaining are displayed throughout the game.

Score keeping: Assigning a score for each successful output produced during the game is the most direct methods of motivating players. The scoring summary differs in each and every game. The scoring pattern adopted in the land cover validation game is quite interesting, since the two different score boards are displayed. First is the “Leader board” which displays the score of the top three players’ position and the player’s current position. Second is the “Best last Players” which shows the best players list according to the cumulative score achieved during the last ten game rounds.

Player Skill level: Another goal-based motivation into GWAP design is by player’s skill levels or ranks. Here the player skill levels are indirectly awarded by assigning different badges. The badges are assigned based on the different task they achieve.

3.5 GWAP Evaluation:

It is necessary to know how enjoyable and effective the game is. In order to know these terms it is necessary to evaluate the game. By doing so, the game's performance can be judged. Ahn and Dabbish (2008)[xlii] describes a set of metrics for determining GWAP success, including throughput, lifetime play, and expected contribution.

It is also necessary to know game efficiency and expected contribution. If we consider games as algorithms, efficiency would be a natural metric of evaluation. There are different algorithms for a given problem and some can be more efficient than other. In the same way there are many GWAP available which are designed to serve for uncertainty problem or which may be efficient for uncertainty problem. In order to choose the best solution to a problem, one has to compare and contrast the different GWAP in terms of efficiency.

The efficiency can defined using throughput which is given by average number of problem instances solved, or input-output mappings performed per human-hours. Learning curves and variation in player skill must be considered in calculating throughput. It’s better to choose the games with higher throughput over the lower throughput. But it’s not the end of the story, because in the GWAP, “fun” also plays the major role, no matter how many problem instances are addressed by a given game, what if nobody wants to play. The real output of utility is given by the combination of throughput and enjoying ability.

Enjoyability is a critical term to quantify and depends on the precise implementation and design of each game. It is vital to know enjoyability of the game. The Average Lifetime Play (ALP) for a game gives the enjoyability of the game. Average Lifetime Play (ALP) is the overall amount of time the game is played by each player averaged across all people who have played it.

Expected contribution is the concise measure of GWAP quality. If the throughput and ALP are known, these metrics can be combined to assess each player's expected contribution. It denotes the average number of problem instances a single human player can be expected to solve by playing a particular game. The programmers use this metrics to evaluate the

GWAPs.

Throughput is the average number of problem instances solved per human-hour which is given by the equation 1

$$Throughput = \frac{Solved\,\mathrm{Pr}\,oblems}{PlayedTime}$$

Equation 1: Throughput equation.

Average Lifetime play is the average (across all people who play the game) overall amount of time that the game will be played by an individual player. It is given by the equation 2.

$$ALP = \frac{PlayedTime}{ActivePlayers}$$

Equation 2: Average Lifetime play equation.

Expected Contribution is the function of throughput and ALP, which is given by the equation 3.

$$\text{Expected Contribution} = Throughput * ALP$$

Equation 3: Expected Contribution equation.

Though we have these metric to capture the efficiency of the game but there are some more aspects such as “popularity” which are much more needed to measure the stability of the game. Generally game enjoyment and fun measurement has been done by using the self-report questionnaire measure. Nevertheless the behavioural measure can be provided directly by assessing the how many people play the game and, in turn, how useful the game is to solve the problem. Lastly, a GWAP's must cross-check that the game's design is indeed correct; that he/she final results should outline properly to the feed input. This can be done by employing the volunteers who can analyses the output.

Chapter 4

4. Land Cover

4.1 Introduction

Several freely available global land cover maps are produced in the recent days with advancement in remote sensing sensors and mapping technologies. The large volume of high-quality global remote-sensed data is the major source of these global land cover maps. In many applications, these remotely-sensed global land cover maps are used without knowing their quality or accuracy. Though these data are freely available, it is necessary to know whether the land cover classification is suitable for the different applications.

The quality of land cover maps depends on the input data and the classification algorithm used to produce them, as well as spatial resolution and legend (Townshend, et al., 1991)[xlix]. The techniques for determining the quality of a particular map is defined as validation. The techniques such as accuracy assessment (overall accuracy, errors of omission, errors of commission by land cover class, allocation disagreement and quantity disagreement), errors analysed by region, and fuzzy accuracy observation are termed as validation techniques.

This chapter describes in details the open datasets which are available. Though there are more dataset which are freely available, in this research work we are focused on the two dataset GlobeLand30 (GL30) and Destinazione d'Uso dei Suoli Agricoli e Forestali (DUSAF). This chapter also describes how these two data are compared to assess the accuracy and to obtain disagreement maps between them to feed into the game.

4.2 Methods and Data

4.2.1 Available data

GlobeLand30 (GL30) which is produced by the Chinese government, then donated to the United Nations for utilizing it for research on sustainable development & climate change and DUSAF dataset, the Italian acronym for "Use Categories of Agricultural and Forest Soil" (Credali et al., 2011)[l] which is the land cover data from the Lombardy region at scale 1:10'000. These data have been made available for the city of Como. The figure 7 shows Regions of Italy with highlighted Lombardy region within it lays the city of Como. The figure 8 shows the Como province with the highlighted Como city. The Como city has been focused for this research work.

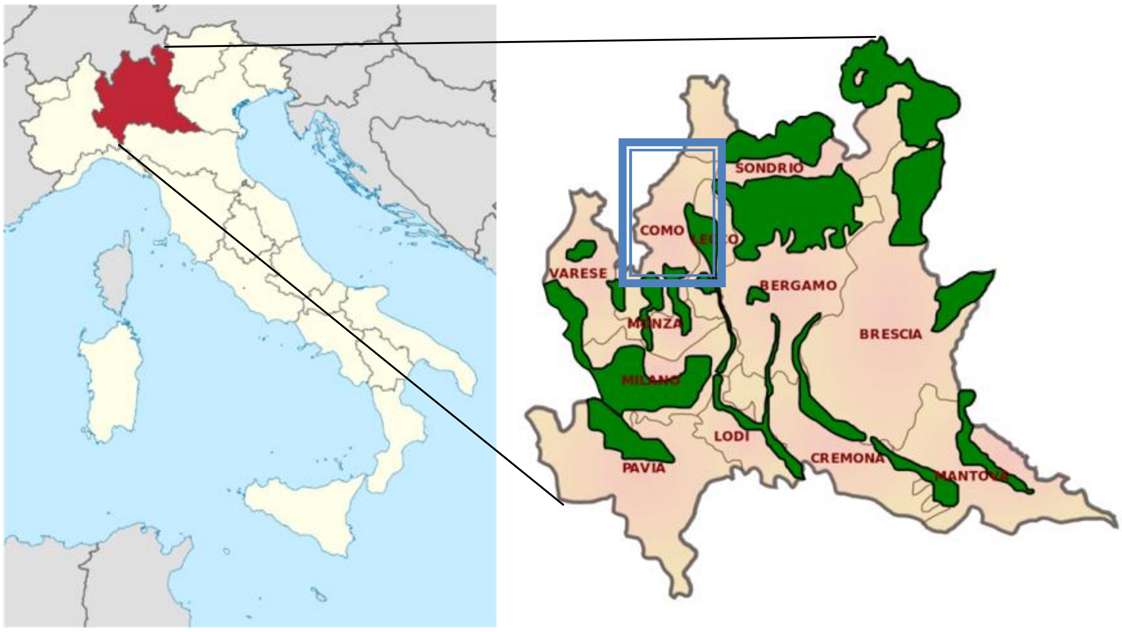


Figure 7 : Regions of Italy with highlighted Lombardy region.

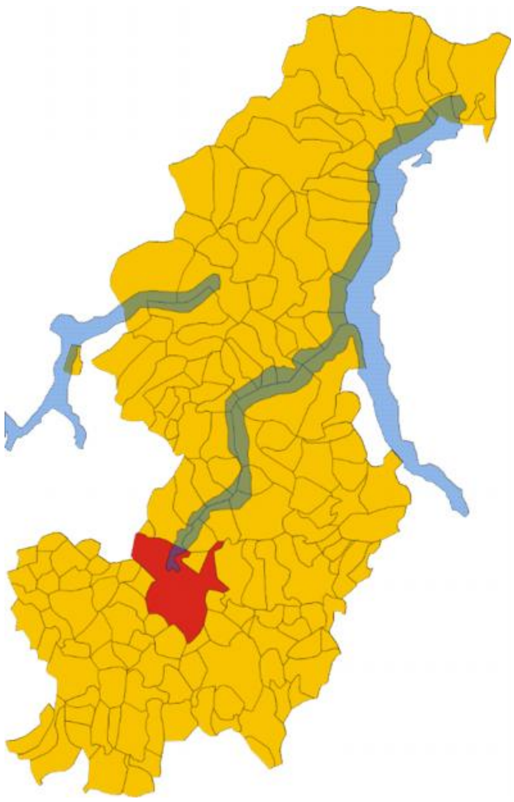


Figure 8: Como province with highlighted Como city.

4.2.1.1 GlobeLand30

GlobeLand30 is the product of "Global Land Cover Mapping at Finer Resolution" project led by the National Geomatics Centre of China (NGCC), which aimed at mapping a global land cover at 30 m resolution. This dataset has been obtained from Landsat Thematic Mapper (TM) and Enhanced TM plus (ETM+) satellites for the baseline 2000. For the baseline 2010 the images of the Chinese Environmental and Disaster (HJ-1) satellite were also used along with the previous satellite sensors. GlobeLand30 has been obtained based on Pixel-Object-knowledge (POK-based) approach (Chen et al., 2014)[li].

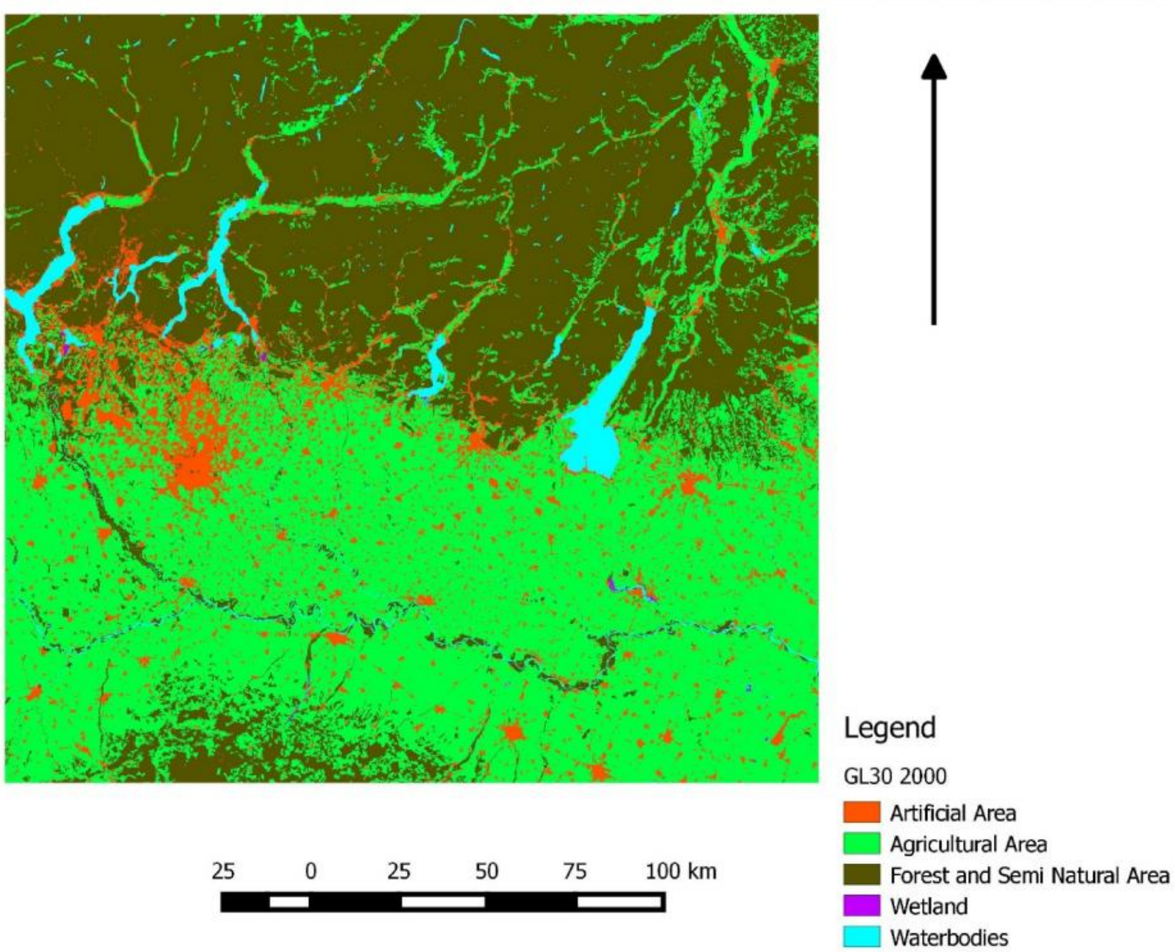


Figure 9: Globe Land 30 of Lombardy region for the baseline year 2000.

For the current study, GL30 is available for the Como municipality which are clipped from the Lombardy GL30 map. The data are available in raster format with World Geodetic System 1984(WGS84) reference system and Universal Transverse Mercator (UTM) projection. The figure 9 shows the GL30 map of Lombardy region for the baseline year 2000. The figure 10 shows the GL30 map of Lombardy region for the baseline year 2010.

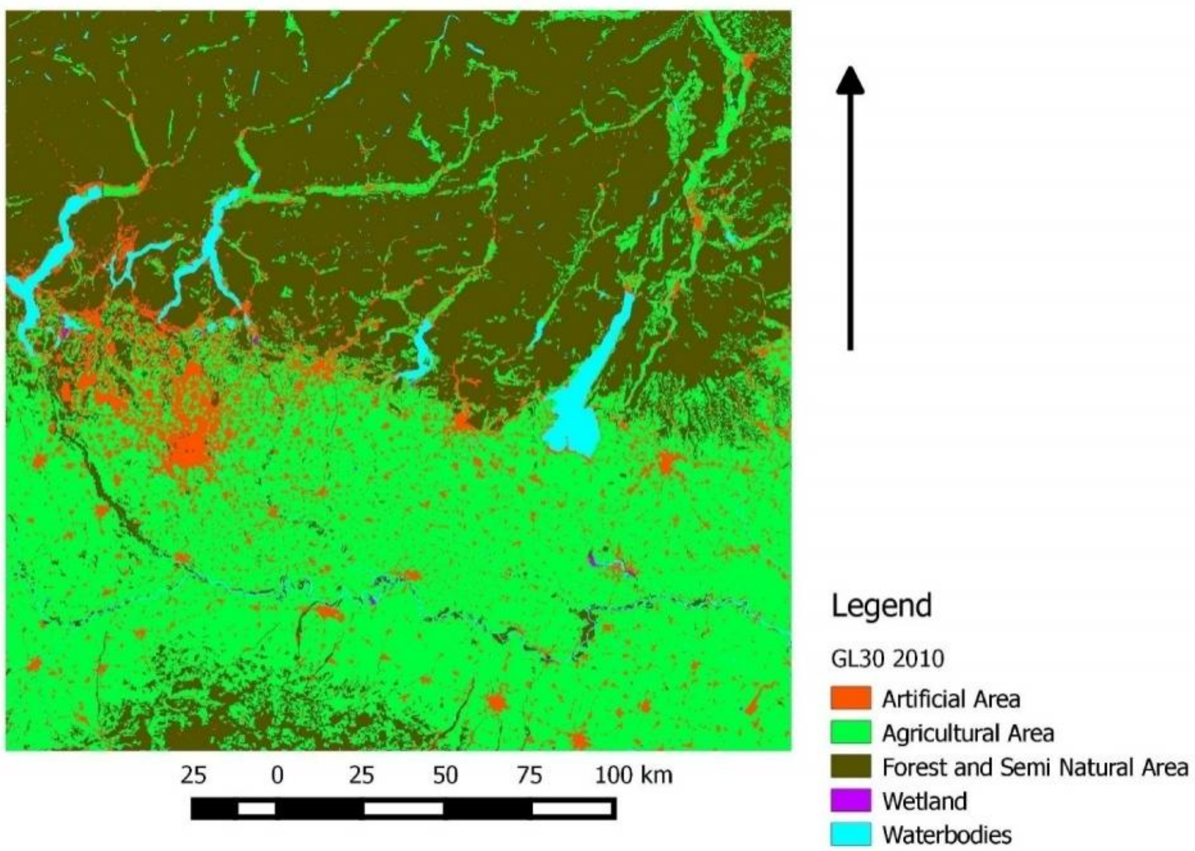


Figure 10: Globe Land 30 of Lombardy region for the baseline year 2010.

4.2.1.2 Italian land cover

DUSAF is a detailed geographical database created by Lombardy Region's Department for Territorial Planning, Agriculture and Forests (ERSAF) with the cooperation of the Lombardy Regional Agency for the Protection of the Environment (ARPA) in 2000-2001 (Credali et al., 2011)[l].

Five releases of the databases are currently available (Table 4)

- **DUSAF 1.0** was prepared from the photo interpretation of aerial photos taken in 1998-1999 (Flight IT2000 made by Blom CGR) using first hierarchical level of 10 classes legend. Later it has been reclassified according to "CORINE Land Cover project" legend nomenclature and named as **DUSAF1.1.**

- **DUSAF 2.0** was prepared from photo interpretation of aerial photos acquired during different period (2005, 2006 and 2007). **DUSAF 2.1** was prepared for the whole region during the period 2007.

- **DUSAF 3.0** was prepared from the acquired aerial photos in 2009 and limited to some parts of Lombardy region.

- **DUSAF 4.0** was prepared for the year 2012.

	DUSAF 1.1	DUSAF 2.0	DUSAF 2.1	DUSAF 3.0	DUSAF 4.0
YEAR	1999 - 2000	2005 - 2007	2007	2009	2012
SCALE	1:10'000	1:10'000	1:10'000	1:10'000	1:10'000
REF SYS	WGS84/UTM32N	WGS84/UTM32N	WGS84/UTM32N	WGS84/UTM32N	WGS84/UTM32N
LEGEND	CORINE	CORINE	CORINE	CORINE	CORINE
SOURCE	Aerial photos	Aerial photos and regional databases	Aerial photos and regional databases	Aerial photos and regional databases	Aerial photos and regional databases
EXTENT	whole region	whole region	whole region	BS, MI, MB, SO, CR	whole region

Table 4: Characteristics of DUSAF database releases.

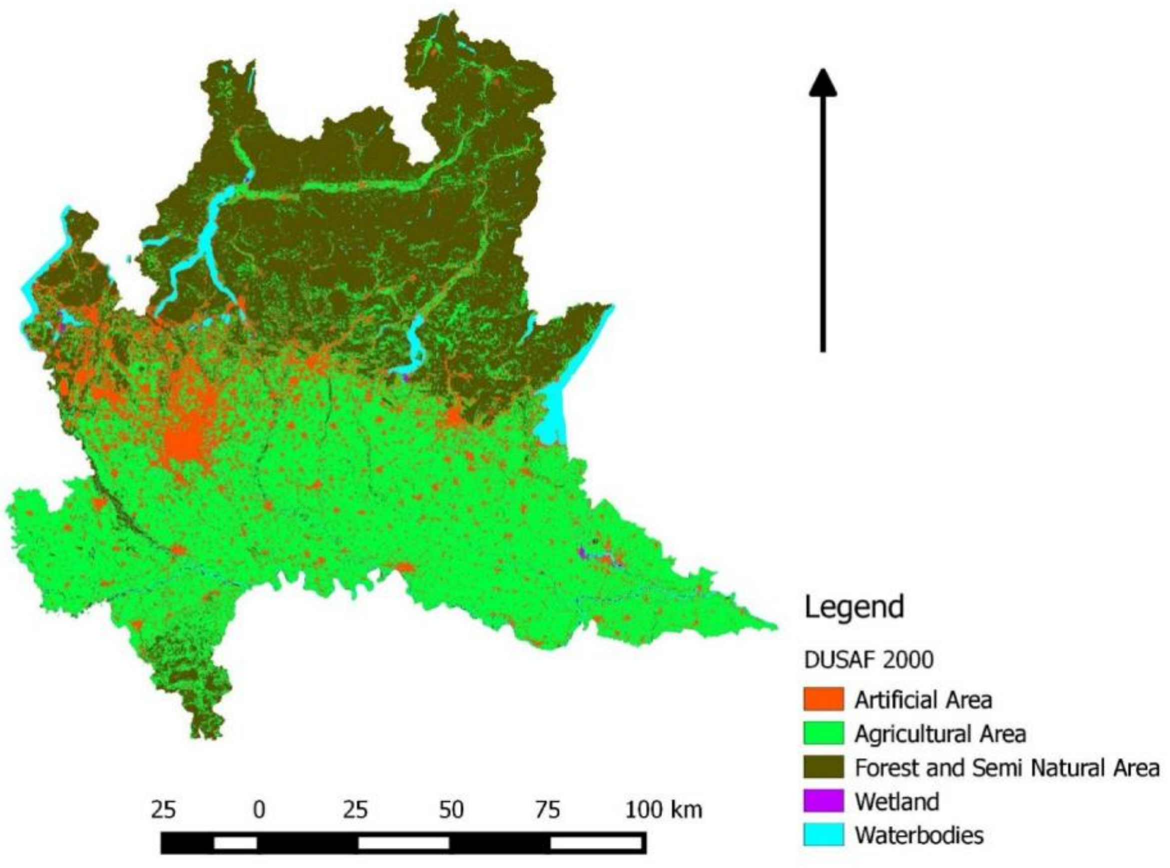


Figure 11: DUSAF 1.1 of Lombardy region for the baseline year 2000.

These datasets are available in ESRI (Environmental Systems Research Institute) shape file format. They are available for different years (Table 4), different spatial reference systems[9], and even different sources including Quick Bird or aerial photo interpretation. The scales[10] of these datasets range between 1:10,000 and 1:25,000 with the accuracies better than 5m. The current DUSAF classification is structured into first level Coordination of Information on the Environment (CORINE) land cover nomenclature, which includes Artificial areas, Agricultural areas, Forest and semi-natural areas, Wetlands and Water bodies.

For the current research, we used DUSAF 1.1 and DUSAF 4.0. The figure 11 shows DUSAF 1.1 for the Lombardy region for the baseline year 2000 and the figure 12 shows DUSAF 4.0 for the Lombardy region which corresponds to the baseline year 2010.

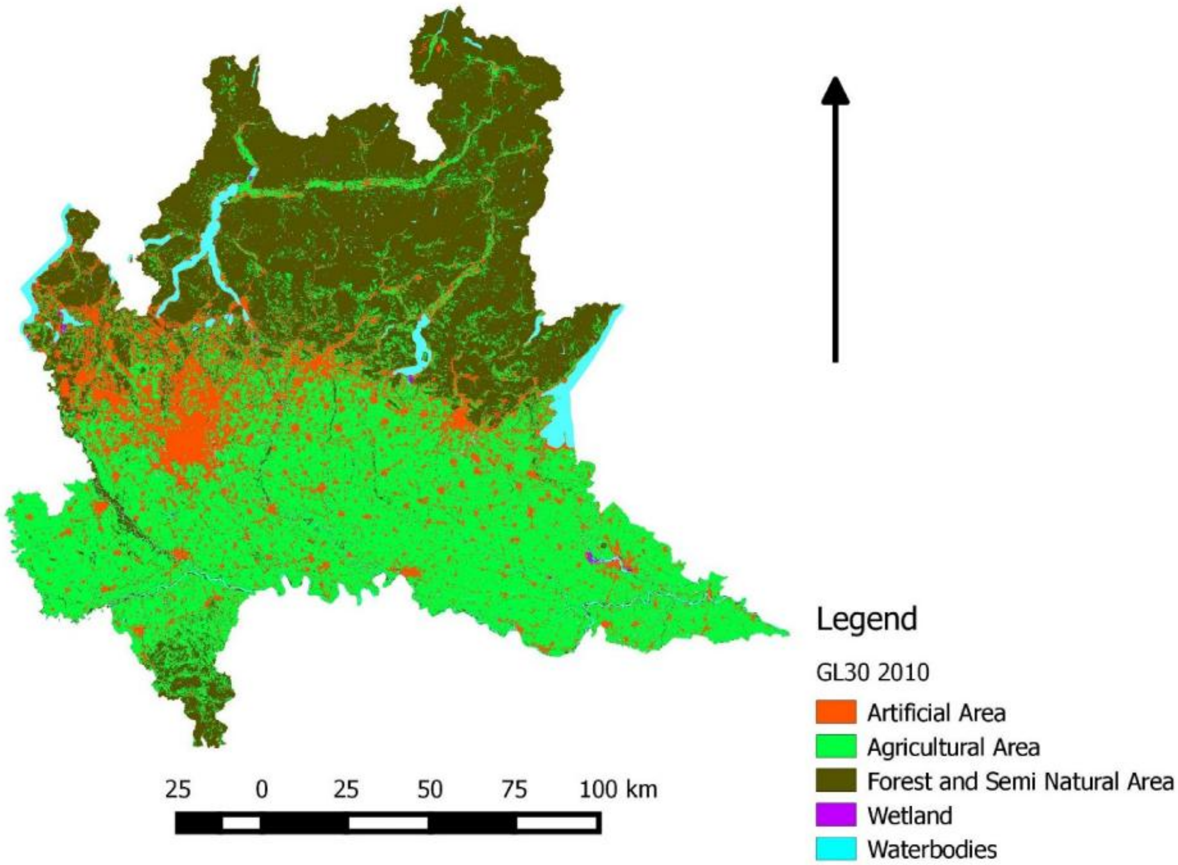


Figure 12: DUSAF 4.0 of Lombardy region for the baseline year 2010.

[9] *A spatial reference system is a coordinate-based local, regional or global system used to locate geographical entities. It defines a specific map projection, as well as transformations between different spatial reference systems.*

[10] *The scale of a map is the ratio of a distance on the map to the corresponding distance on the ground.*

4.2.2 Validation

4.2.2.1 Accuracy assessment

The main purpose of accuracy assessment is to obtain a quantitative description of the accuracy of the global land cover map. The basis of the accuracy assessment is to compare the classes of the data with the ground truth. Though it is challenging in terms of time and money spent. There are many methods and indexes used to measure the classification quality of thematic maps. Here we adopt the accuracy analysis carried out by means of the confusion or error matrix, which has been suggested as "Good Practices" by Foody (2011)[lii]. The confusion matrix is easy to interpret and familiar to both the map user and producer communities.

GROUND TRUTH (REFERENCE)

CLASSIFICATION	A	B	C	D	
A	f_{AA}	f_{AB}	f_{AC}	f_{AD}	f_{A+}
B	f_{BA}	f_{BB}	f_{BC}	f_{BD}	f_{B+}
C	f_{CA}	f_{CB}	f_{CC}	f_{CD}	f_{C+}
D	f_{DA}	f_{DB}	f_{DC}	f_{DD}	f_{D+}
	f_{+A}	f_{+B}	f_{+C}	f_{+D}	n

Figure 13: Design of Confusion or Error matrix.

Here the spatial comparison between the classified dataset and reference one is performed by selecting the pixel as the spatial unit (Broveli et al 2015)[iv]. This comparison results in a square matrix by the number of rows and column equal to the total number of the considered land cover classes as shown in the figure 13. The elements along the diagonal represent the correctly classified data, whereas the off-diagonal elements identify the classification errors. The observation including Overall Accuracy (OA), User Accuracy (UA), Producer Accuracy (PA), Allocation Disagreement (AD) and Quantity Disagreement (QD) have been considered as "Good Practices" for the accuracy assessment.

$$OA = \frac{\sum f_{ii}}{n} (i=A,B,C,D)$$

Equation 4: Overall Accuracy.

Overall accuracy is one of the most popular agreement measures which give the percentage of correctly classified pixels. It can be computed using the equation 4 above, where q is the number of the classes, n_{ii} are the diagonal elements and n represents the total number of considered pixels. There is no literature which explains the threshold for acceptable limit for the overall accuracy for the image classification. Anderson et al. (1976)[liii] propose a value of at least 85%, Pringle et al.(2009)[liv] recommend a value over 70%, whereas Thomlinson et al. (1999)[lv] consider accurate a classification when the OA is at least equal to 85% and no class is less than 70%.

$$UA_i = \frac{f_{ii}}{f_{i+}} \quad (i=A,B,C,D)$$

Equation 5: User Accuracy.

$$CE_i = 1 - UA_i \quad (i=A,B,C,D)$$

*Equation 6**:** Commission Error.*

The User accuracy (Equation 5) explains the accuracy of the classification from the user perspective and is defined as the percentage of the classified pixels that exactly match the ground truth, whereas the producer accuracy (Equation 7) is defined as the percentage of the pixels of ground truth correctly detected in the classified map. The errors such as errors of commission (Equation 6) and errors of omission (Equation 8) correspond to User's and Producer's respectively.

$$PA_i = \frac{f_{ii}}{f_{+i}} \qquad (i=A,B,C,D)$$

*Equation 7**:** Producer Accuracy.*

$$OE_i = 1 - PA_i \quad (i=A,B,C,D)$$

*Equation 8**:** Omission Error.*

Kappa is the proportion of agreement after chance agreement is removed (Rosenfield and Fitzpatrick-Lins, 1986)[lvi]. From the error matrix, Kappa coefficient is calculated using the equation 9.

$$k = \frac{\frac{1}{n}\sum f_{ii} - \frac{1}{n^2}\sum f_{i+}f_{+i}}{1 - \frac{1}{n^2}\sum_i f_{i+}f_{+i}}$$

$$k_i = \frac{\frac{f_{ii}}{n} - \left(\frac{f_{i+}}{n} * \frac{f_{+i}}{n}\right)}{\frac{1}{2}\left(\frac{f_{i+}}{n} + \frac{f_{+i}}{n}\right) - \left(\frac{f_{i+}}{n} * \frac{f_{+i}}{n}\right)} (i = A, B, C, D)$$

Equation 9: Kappa Standard.

The value of Kappa is between 1 and –1, the higher the value, the stronger the agreement. Although the kappa coefficient has been widely promoted for accuracy assessment (*e.g.,* Congalton *et al.,* 1983)[lvii], there are sufficient concerns with its use (*e.g.,* Foody, 1992[lviii]; Stehman and Czaplewski, 1998)[lix] that it cannot be recommended as a general measure of map accuracy. Foody (2008)[xii] exposed some of the conceptual problems with the standard Kappa; the arguments used to promote the use of the kappa coefficient are fundamentally flawed.

$$AD_i = 2 * min\left(\frac{f_{+i}}{n} - \frac{f_{ii}}{n}, \frac{f_{i+}}{n} - \frac{f_{ii}}{n}\right) \qquad i = (A, B, C, D)$$

$$AD = \frac{\sum AD_i}{2} \qquad i = (A, B, C, D)$$

Equation 10: Allocation Disagreement.

The more recent approach processed by Pontius and Millones (2008)[lx] estimates the disagreement component between classified and reference datasets. The allocation disagreement (Equation 10) is defined as the disagreement value that is due to less optimal match in the spatial allocation of the categories, while the quantity disagreement (Equation 11) is due to the less than perfect match in the proportion of the categories.

$$QD_i = \left|\frac{f_{+i}}{n} - \frac{f_{i+}}{n}\right| \qquad i = (A, B, C, D)$$

$$QD - \frac{\sum QD_i}{2} \qquad i = (A, B, C, D)$$

Equation 11: Quantity Disagreement.

4.2.2.2 Data Processing

Since GlobeLand30 and Italian land cover maps have different characteristics such as format, legend, scale and reference system, it is necessary to pre-process the data before computing the confusion matrix and its derived statistics. The data processing work flow (Figure 14) shows that Italian land cover has been converted into raster before reclassification since they are provided as the vector datasets.

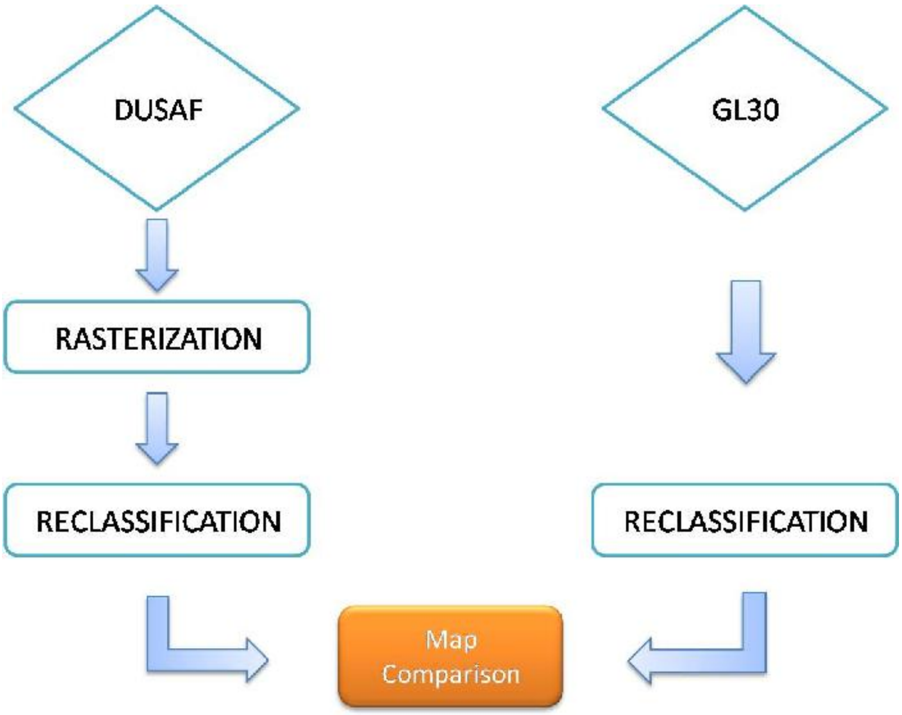


Figure 14: Data processing work flow.

CORINE LEGEND	
1. Artificial cover	
2. Cropland	
3. Forest and semi natural areas	31. Forest
	32. Scrub and/or herbaceous vegetation associations
	33. Beaches, dunes, sands, bare rocks, sparsely vegetated areas
	335. Glaciers and perpetual snow
4. Wetland	
5. Open Water	

Table 5: CORINE Land Cover legend.

The GlobeLand30 and the Italian datasets have a different thematic classification; hence the reclassification process is carried out to classify both datasets based on the first five levels of CLC (CORINE Land Cover) nomenclature (Table 5). The table 6 describes the reclassification of GL30 classes based on the CORINE legend.

Corine Legend	GlobeLand30 Legend
1 Artificial surfaces	1 Artificial cover
2 Agricultural areas	2 Croplands
3 Forests and semi natural areas	3 Mixed forest, Broadleaf forest, Coniferous forest, Grass, Shrub, Bare land, Permanent ice or snow
OR	OR
3.1 Forests	3.1 Mixed forest, Broadleaf forest, Coniferous forest
3.2 Scrub and/or herbaceous vegetation associations	3.2 Grass, Shrub
3.3.0 Open spaces with little or no vegetation	3.3.0 Bare land
3.3.5 Glaciers and perpetual snow	3.3.5 Permanent ice or snow
4 Wetlands	4 Wetlands
5 Water bodies	5 Water

Table 6: Reclassification of GL30 based on CORINE legend.

4.2.2.3 Comparison between GlobeLand30 2000 and DUSAF 1.1

GlobeLand30 2000 and DUSAF 1.1 have been compared for the Como area by making DUSAF1.1 as ground truth and the statistics including overall accuracy, user accuracy, producer accuracy allocation disagreement and quantity disagreement has been obtained. The table 7 shows Confusion matrix and agreement measures between GL30 and DUSAF 1.1 for the Como area. The figure 15 gives GlobeLand30 2000 and DUSAF 1.1 for the Como area.

CLASSIFIED GL30 2000 \ GROUND TRUTH DUSAF1.1	1	2	3	4	5	sum	UA	CE
1	16282	1221	3074	0	95	20672	78.76%	21.24%
2	647	1625	952	4	0	3228	50.34%	49.66%
3	932	902	13154	98	32	15118	87.01%	12.99%
4	0	0	0	0	0	0	-	-
5	54	1	4	6	2357	2422	97.32%	2.68%
sum	17915	3749	17184	108	2484	41440		
PA	90.88%	43.34%	76.55%	0.00%	94.89%			
OE	9.12%	56.66%	23.45%	100.00%	5.11%			

OA	K	AD	QD
80.64%	0.689034	12.71%	6.65%

Table 7: Confusion matrix and agreement measure between GL30 2000 and DUSAF 1.1.

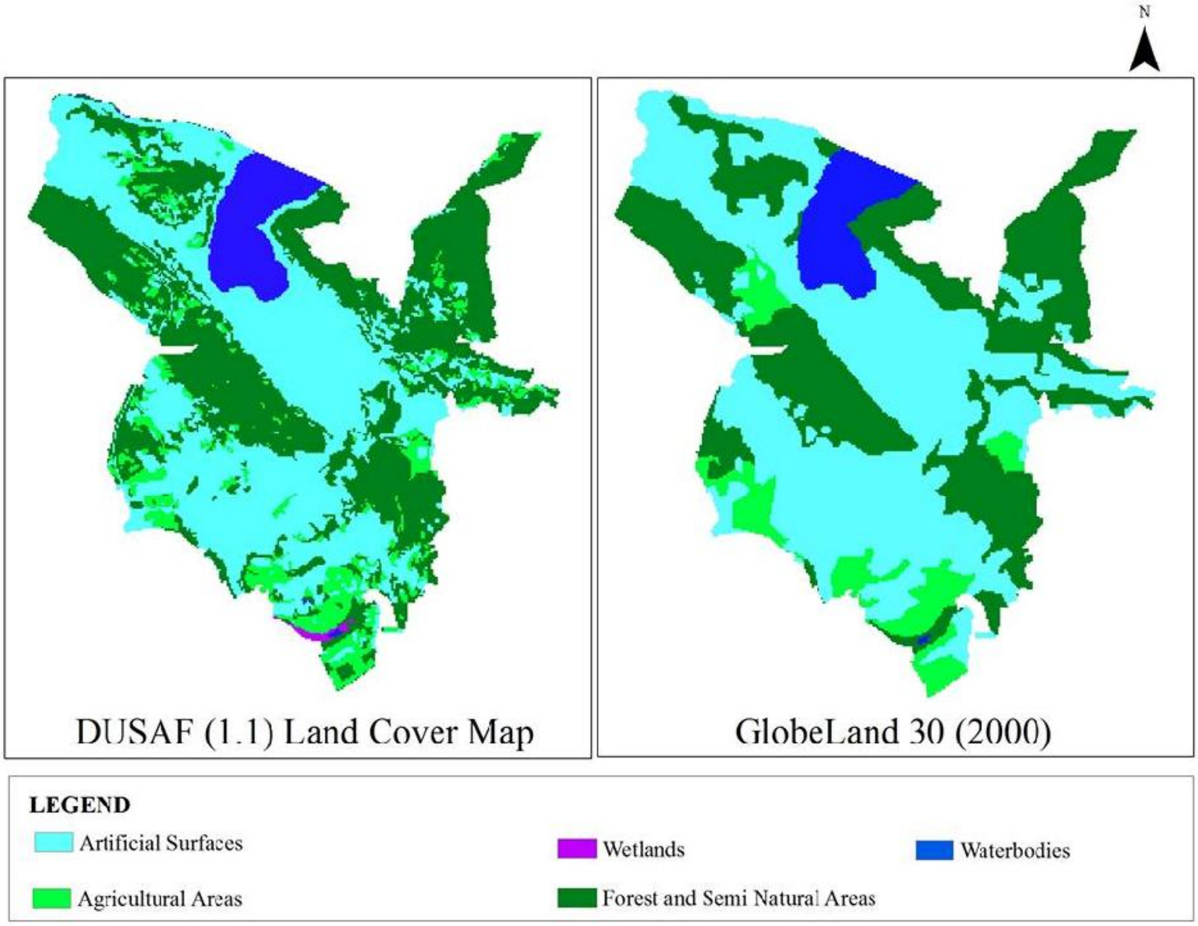


Figure 15: GlobeLand30 2000 and DUSAF 1.1 for the Como area.

4.2.2.4 Comparison between GlobeLand30 2000 and DUSAF 4.0

GlobeLand30 2010 and DUSAF 4.0 have been compared for the Como area by making DUSAF 4.0 as the ground truth[11]. The table 8 shows Confusion matrix and agreement measures between GL30 2010 and DUSAF 4.0 for the Como area which includes the overall accuracy, user accuracy, producer accuracy allocation disagreement and quantity disagreement. Though DUSAF 4.0 corresponds to the year 2012, we decided to use it because of the incompleteness of DUSAF 3.0. Also the evaluation of differences between DUSAF 3.0 and DUSAF 4.0 in their overlapping area shows an agreement of 99.8% between them (Brovelli et al 2015)[iv]. The figure 16 gives GlobeLand30 2010 and DUSAF 4.0 for the Como area.

[11] *Ground truth is a term used in various fields to refer to the absolute truth of something.*

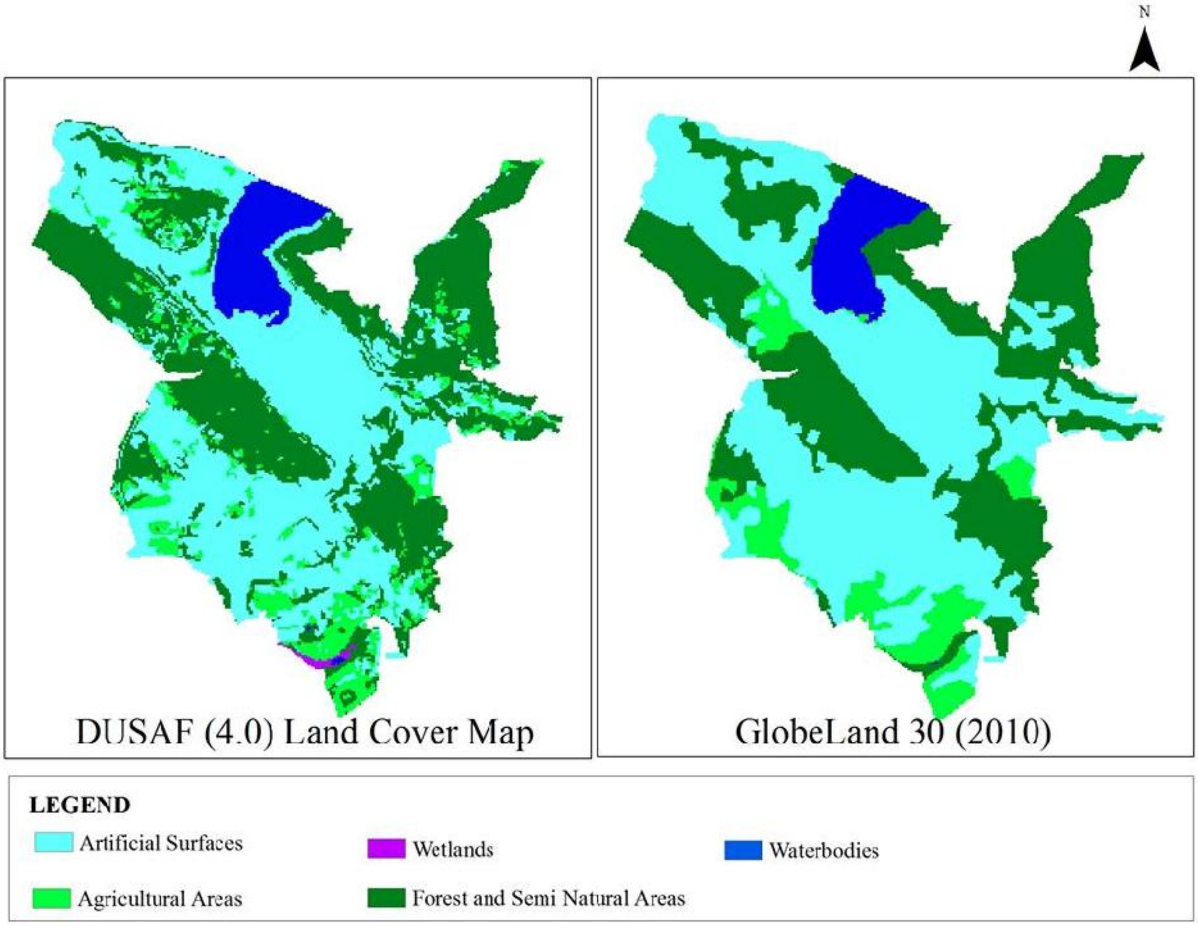


Figure 16: GlobeLand30 2010 and DUSAF 4.0 for the Como area.

CLASSIFIED GL30 2010	GROUND TRUTH DUSAF 4.0: 1	2	3	4	5	sum	UA	CE
1	16629	1340	2626	0	131	20726	80.23%	19.77%
2	917	1644	872	29	9	3471	47.36%	52.64%
3	1177	835	12824	79	46	14961	85.72%	14.28%
4	0	0	0	0	0	0	0	0
5	57	0	2	0	2221	2280	97.41%	2.59%
sum	18780	3819	16324	108	2407	41438		
PA	88.55%	43.05%	78.56%	0.00%	92.27%			
OE	11.45%	56.95%	21.44%	100.00%	7.73%			

OA	K	AD	QD
80.40%	0.684032	14.90%	4.70%

Table 8: Confusion matrix and agreement measure between GL30 2010 and DUSAF 4.0.

4.3 Data preparation for the validation game

The GlobeLand30 and DUSAF land cover data products have been compared and difference map has been produced by subtracting the GL30 and DUSAF land cover maps. In order to extract the pixel coordinates (as vector) along with the GL30 and DUSAF classification to feed into the Land Cover Validation game, some of the pre-processing steps have been performed. The figure 17 shows the difference map of GL30 and DUSAF for the year 2000 and figure 18 shows the difference map of GL30 and DUSAF for the year 2010.

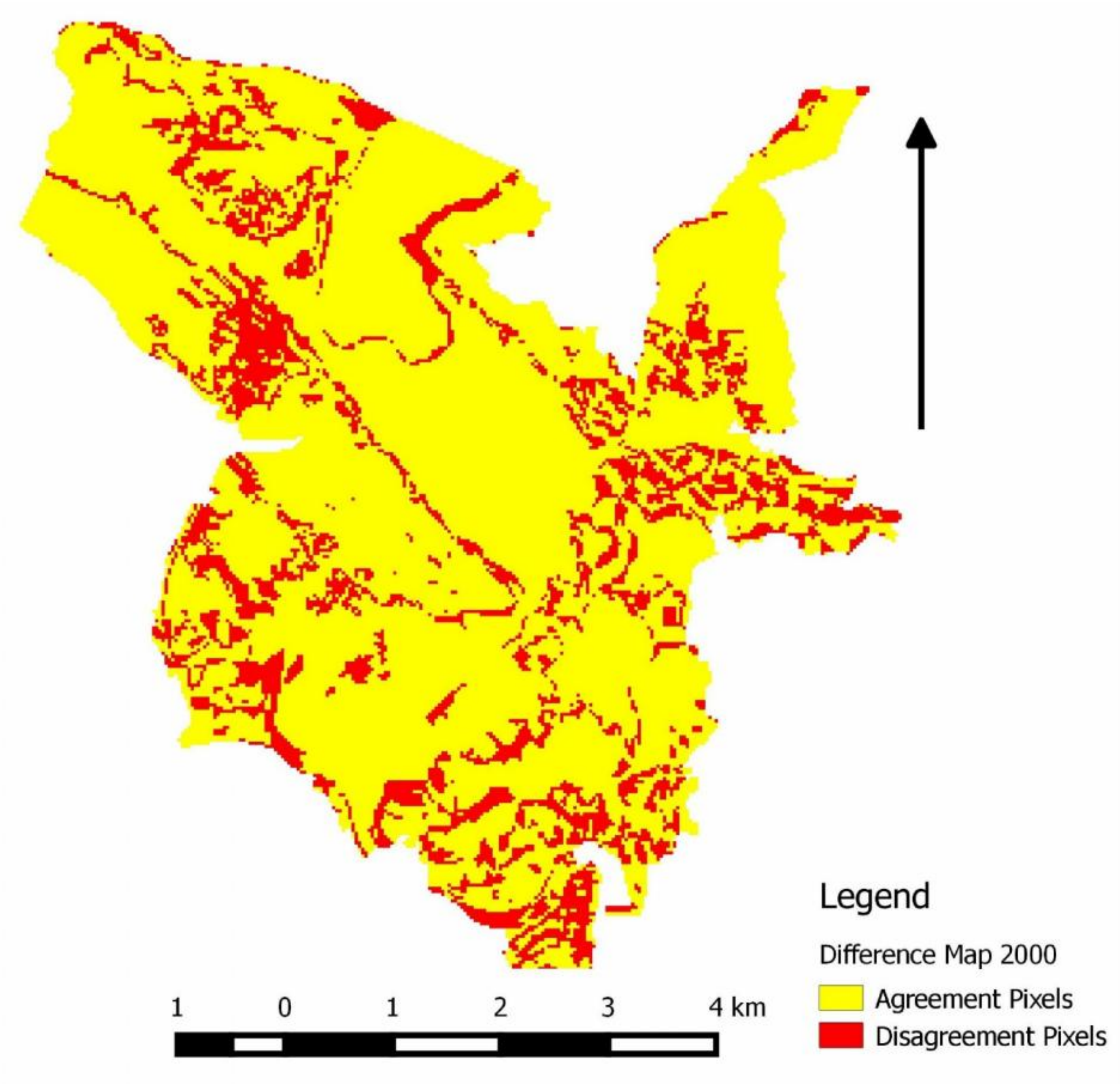


Figure 17: Difference Map of GL30 and DUSAF for the baseline year 2000.

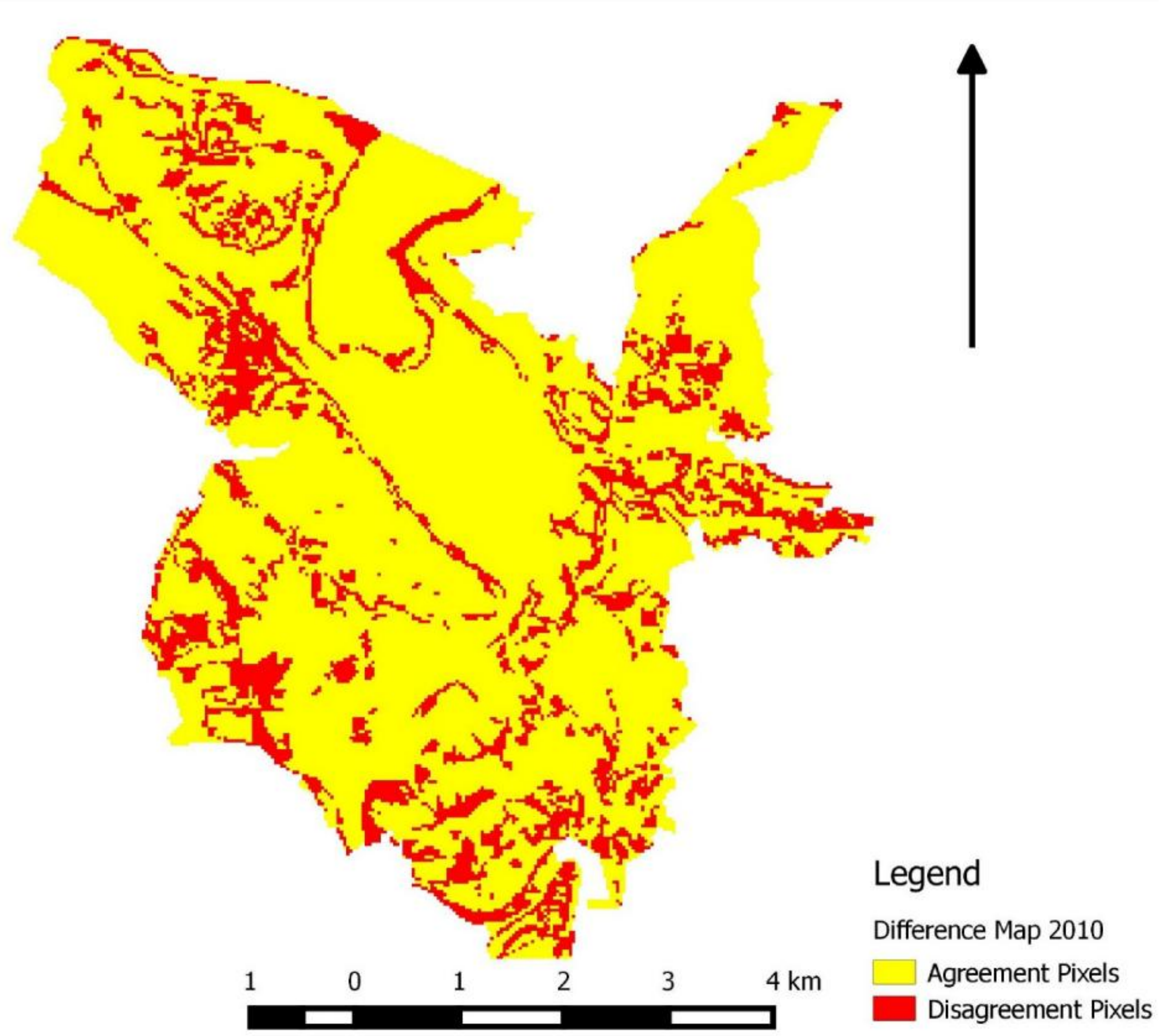


Figure 18: Difference Map of GL30 and DUSAF for the baseline year 2010.

The workflow diagram of this process is given in the figure 19. Initially the grid of 30 m for Como region is prepared using GRASS software. By computing the grid for the Como region it is possible to get the spatial extent of each pixel. Later the centroid is computed for this grid. Using SAGA tool[12] "Add polygon value to points", Difference map attribute values are added to the point vector data (grid centroid). This pixel coordinates along with GL30 and DUSAF classification has been saved as CSV file (Figure 20) for uploading into the game. In the figure 20 the first column is id, second and third column (row & column) corresponds to the grid of the pixels. DIFF2010 column denotes whether GL30 and DUSAF agree or not, value 0 denotes agreement between them and 1 denotes the disagreement. The DUSAF and GL30 column represents the classification for each pixel. The X and Y column gives the coordinates for each pixel.

[12] *SAGA tool is a module library that contains various tools to manipulate raster datasets (including stud like changing cell values, reclassification, buffering, merging patching, resampling, etc)*

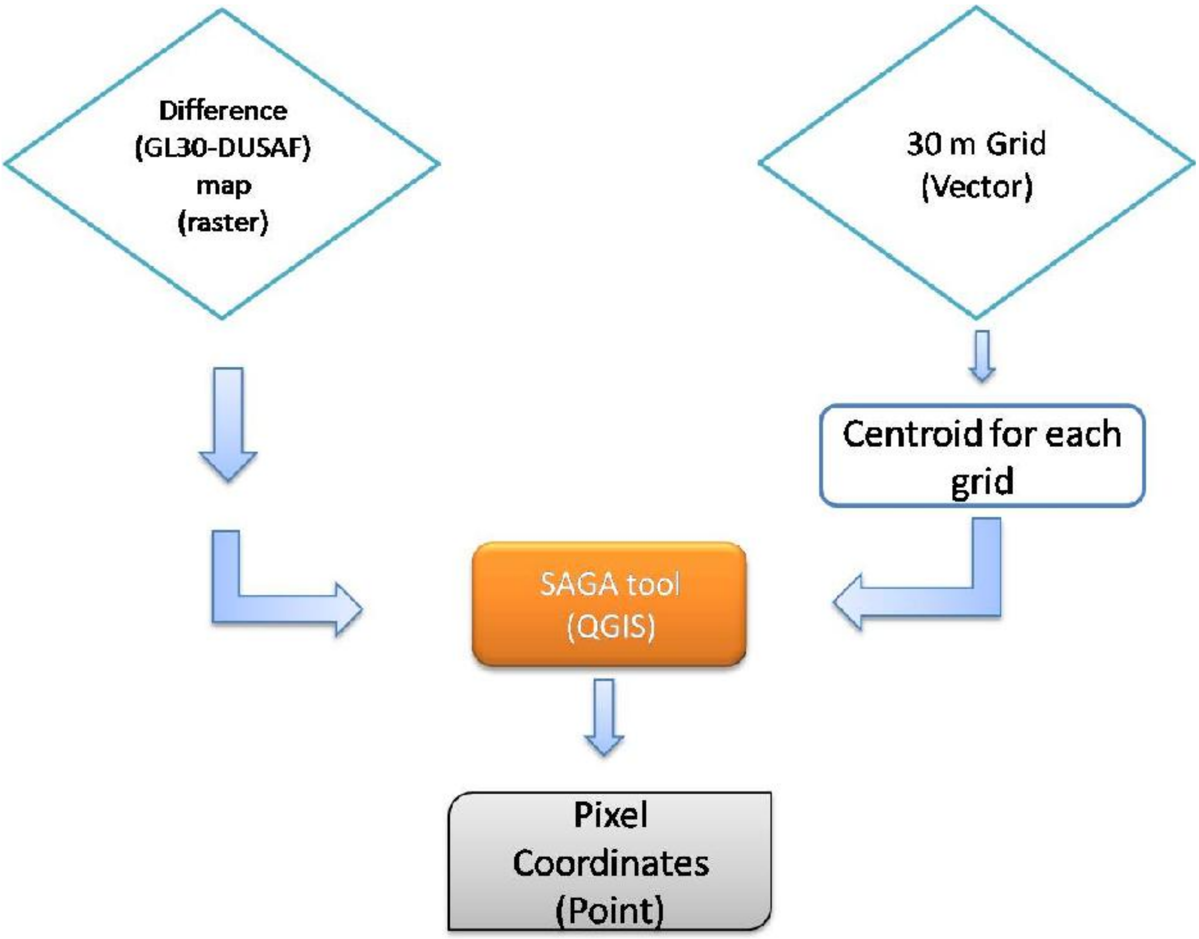
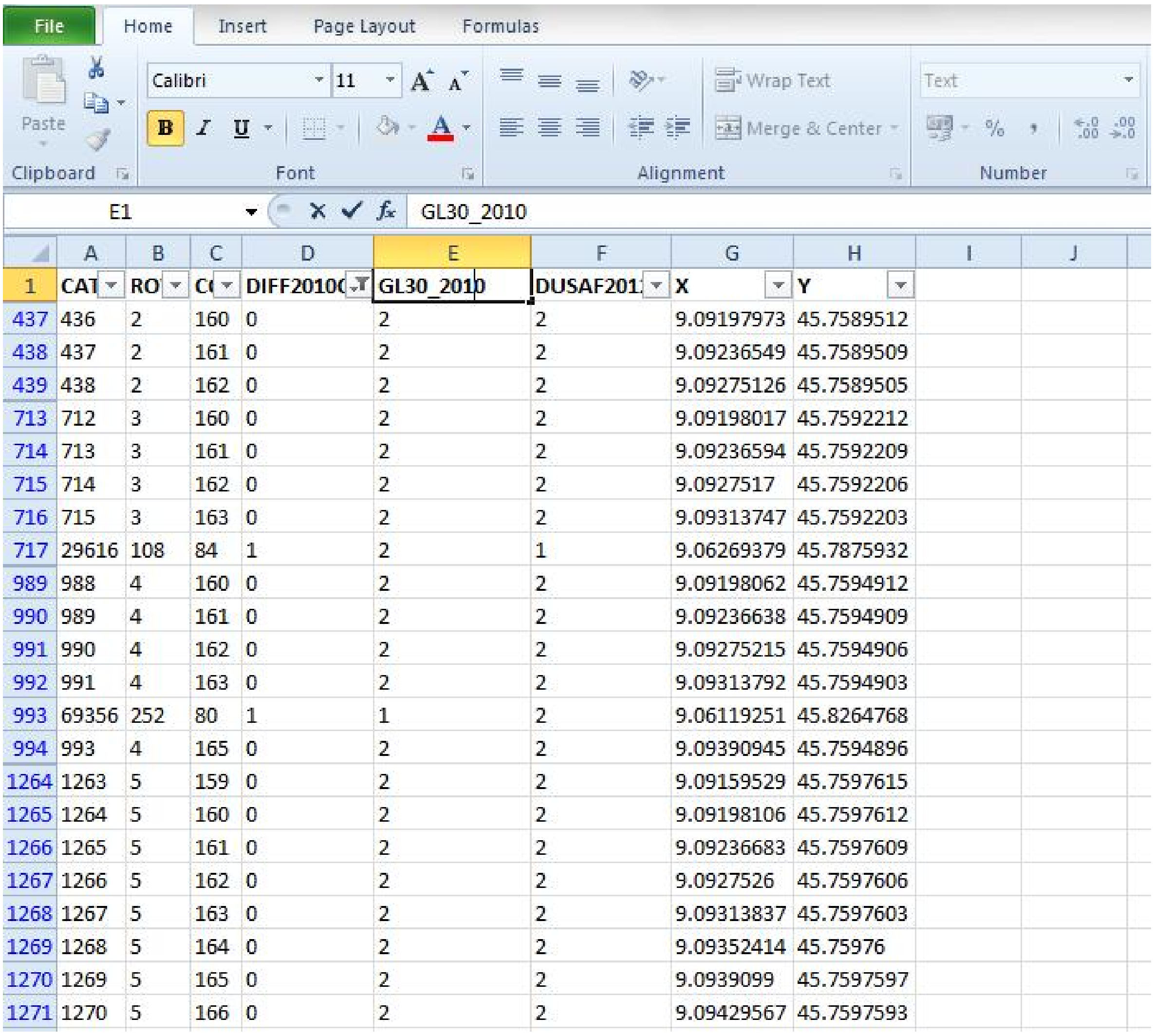


Figure 19: Workflow diagram to extract disagreement pixels.



	A	B	C	D	E	F	G	H
1	CAT	RO	C(	DIFF2010(	GL30_2010	DUSAF201:	X	Y
437	436	2	160	0	2	2	9.09197973	45.7589512
438	437	2	161	0	2	2	9.09236549	45.7589509
439	438	2	162	0	2	2	9.09275126	45.7589505
713	712	3	160	0	2	2	9.09198017	45.7592212
714	713	3	161	0	2	2	9.09236594	45.7592209
715	714	3	162	0	2	2	9.0927517	45.7592206
716	715	3	163	0	2	2	9.09313747	45.7592203
717	29616	108	84	1	2	1	9.06269379	45.7875932
989	988	4	160	0	2	2	9.09198062	45.7594912
990	989	4	161	0	2	2	9.09236638	45.7594909
991	990	4	162	0	2	2	9.09275215	45.7594906
992	991	4	163	0	2	2	9.09313792	45.7594903
993	69356	252	80	1	1	2	9.06119251	45.8264768
994	993	4	165	0	2	2	9.09390945	45.7594896
1264	1263	5	159	0	2	2	9.09159529	45.7597615
1265	1264	5	160	0	2	2	9.09198106	45.7597612
1266	1265	5	161	0	2	2	9.09236683	45.7597609
1267	1266	5	162	0	2	2	9.0927526	45.7597606
1268	1267	5	163	0	2	2	9.09313837	45.7597603
1269	1268	5	164	0	2	2	9.09352414	45.75976
1270	1269	5	165	0	2	2	9.0939099	45.7597597
1271	1270	5	166	0	2	2	9.09429567	45.7597593

Figure 20: CSV output of the difference map.

4.4 Determining the sample size

The previous analysis resulted in the following number of Coherent and Non-Coherent pixels (Table 9). But validating more than 8000 pixels is tedious and time consuming process. The validation process may be boring if Citizen Scientist has to verify 8000 pixels. In order to make validation effective the sampling design has been introduced. The sampling design is the protocol for selecting the subset of spatial units (e.g., pixels or polygons) that will form the basis of the accuracy assessment. Choosing a sampling design requires a consideration of the specific objectives of the accuracy assessment and a prioritized list of desirable design criteria. The most critical recommendation is that the sampling design should be a probability sampling design. Probability sampling is defined in terms of inclusion probabilities, where an inclusion probability relates the likelihood of a given unit being included in the sample (Stehman, 2000)[lxi]. The major decisions in choosing a sampling design relate to tradeoffs among different designs in terms of advantages to meet specified accuracy objectives and priority desirable design criteria. So the desired figure to validate the pixel has been obtained from Cochran's equation. Cochran, (1977)[lxii] suggests using a sample size of n (Equation 12) for simple random sampling and targeting overall accuracy as the estimation objective.

$$n = \frac{z^2 O(1-O)}{d^2}$$

Equation 12: Sample size for simple random sampling.

Where O is the overall accuracy expressed as a proportion, z is a percentile from the standard normal distribution (z =1.96 for a 95% confidence interval, z = 1.645 for a 90% confidence interval), and d is the desired half-width of the confidence interval of O.

	Coherent pixels	**Non-coherent pixels**	**Overall Accuracy [%]**	**Disagreement [%]**
2000	**33416**	**8022**	**80.60**	**19.40**
2010	**33318**	**8120**	**80.40**	**19.60**

Table 9: Count of Coherent and Non-Coherent pixels for 2000 and 2010.

For stratified random sampling, Cochran provides the following sample size as in the equation 13.

$$n = \frac{\left(\sum W_i S_i\right)^2}{\left[S(O)\right]^2 + (1/N)\sum W_i S_i^2}$$

Equation 13: Sample size for stratified random sampling.

where N is the number of units in the Region Of Interest (ROI), S(O) is the standard error of estimated overall accuracy that we would like to achieve, W_i is the mapped proportion of area of class i, S_i is the standard deviation of stratum i & it is computed from user accuracy. Cochran, (1977)[lxii] specify a target standard error for overall accuracy of 0.01.

	Ui	Si	Wi	Si^2	Si*Wi	$Wi*Si^2$
1	0.7876	0.409006	0.432312	0.167286	0.176818	0.07232
2	0.5034	0.499988	0.090468	0.249988	0.045233	0.022616
3	0.8701	0.336193	0.414672	0.113026	0.13941	0.046869
4	0.3956	0.488979	0.002606	0.239101	0.001274	0.000623
5	0.9732	0.161498	0.059942	0.026082	0.009681	0.001563
					0.372416	0.143991

n
1340

Table 10: Sample size computation of the Como for the year 2000.

The table 10 shows the sample size computation of the stratified random sampling using the equation 13 for the year 2000 and table 11 shows the sample size computation of the stratified random sampling using the equation 13 for the year 2010. The number of pixels to be validated for the year 2000 and 2010 is 1340 and 1208.

	Ui	Si	Wi	Si2	Si*Wi	Wi*Si2
1	0.8003	0.399775	0.453207	0.15982	0.181181	0.072432
2	0.8518	0.355298	0.092162	0.126237	0.032745	0.011634
3	0.8925	0.309748	0.393938	0.095944	0.122021	0.037796
4	0.6191	0.485608	0.002606	0.235815	0.001266	0.000615
5	0.9207	0.270206	0.058087	0.073012	0.015695	0.004241
					0.352908	0.126717

n
1208

Table 11: Sample size computation of the Como for the year 2010.

The validation process intended to validate 1340 pixels to assess the quality of the produced land cover map for the year 2000 and 1208 pixels have to be validated to evaluate the quality of the produced land cover map for the year 2010. These sizes of pixels are much easier for the Citizen Scientist to classify without losing motivation when compared with more than 8000 pixel validation.

Chapter 5

5. Land Cover Validation Game

5.1 Introduction

Land Cover Validation game is an uncomplicated game in which each player is displayed with a blue square box (pixel of 30m resolution) which is placed above the high resolution orthophoto (kindly provided by CGR).The pixels correspond to non-coherent pixels (i.e.) which are disagreements between two classifications (DUSAF and GL30). The players are asked to choose an apt category of land cover classification from the five categories displayed for the given pixel. In the five categories displayed one category is the result of DUSAF classification, one category is the result of GL30 classification and the other three categories are the first level CORINE Nomenclature which does not belong to both GL30 and DUSAF classification.

The players scores point if they agree with one of the existing classification (DUSAF or GL30) and doesn't score points if they doesn't agree with one of the existing classification. The players also score more points for consecutive agreements. The land cover data is being validated pixel by pixel based on the maximum agreement for the same classification. The player's reputation is computed based on the errors they make in the game. The classification is marked as error if the player doesn't agree with classification of DUSAF or GL30.The confidence level of pixel increase based on the player's reputation. This helps in increasing the robustness of the land cover validation. The methodology section 5.5.1.3 explains in detail about how the reputation, error and score is computed

The game was initially tested by the experts from the GEO lab[13]and CEFRIEL[14] for the performance. The performance turning was done which has been explained in Appendix A. After testing the game it has been introduced to the crowd in the Mapping party of the Free and Open Source Software for Geospatial (FOSS4G) conference for the validation purpose. Compromising result has been obtained during the three days of the conference. This Land cover validation game can be played online using the link http://landcover.como.polimi.it/landcover/ in any kind of device including laptops, desktop computers, Smartphones or tablets. The incentive for the participants has been awarded by different means. Instant incentives in the form of badges have been awarded within the game for the particular task, which is designed in the game. Apart from that, the top three players have been awarded with the gifts at the end of mapping party (see Chapter 6).

This chapter elaborates the Land Cover Validation game, in particular the components that make up the system, how these components fit together to form a single architecture, and their standards and protocols that have driven this design. It also elaborates about the pre-

[13] *http://geomatica.como.polimi.it/*

[14] *http://www.cefriel.com/en/*

processing of base layers in Geoserver[15], how the validation process is made possible with the Hypertext Pre-processor[16] (PHP) scripting, how the threshold of confidence level is set. The following things will be explained in detail.

5.2 Requirement

This dynamic game, which validates the Land Cover meets some of the requirements for the users as well as device to play the game. The requirements are listed below.

1. The player should be able to access the Land Cover Validation game in any device online, so the basic web browser is needed to play the game.
2. The Land Cover Validation game can be logged in only using any one of the social networking services such as Facebook, Twitter or Google Plus. Therefore the player should have one of the above mentioned social networking service credentials to access the game.
3. For the purpose of authentication, player should allow the Land Cover Validation game to access the basic information from their profile, including their first name, last name and profile picture.
4. Though the fast internet connection is not required, but still it is desirable.
5. The player should possess basic knowledge in land cover classification or at least image interpretation techniques.
6. A user can play one or more rounds
7. Every round has a time limit of 100 sec before which the players have to complete the game.
8. A round is made of five levels, when the five levels are completed the round ends.
9. For each level there is a pixel (resource) and five categories (first level CORINE nomenclature) including
    a. Artificial Surfaces
    b. Agricultural Areas
    c. Forest and Semi Natural Areas
    d. Wetland

---

[15] *GeoServer is an open-source server written in Java allows users to share, process and edit geospatial data.*

[16] *Hypertext Preprocessor is a server-side scripting language designed for web development but also used as general-purpose programming language.*

e. Water bodies

10. Out of these five categories, two categories belong to DUSAF and GL30 classification.

11. At the end of each level (pixel):

    a. If the player chooses DUSAF or GL30 classification a GREEN message is shown “You agree with existing classifications”, otherwise a RED one “You disagree with (both) existing classifications.

    b. The player is assigned with the score as shown below

*Score= Points*100 + N^ of correct consecutive answers * 50*

Equation 14: Score Computation.

Where Points = 50 *if GREEN* (*Selection of DUSAF or GL30 category*)

Points = 0 *if RED (Selection of categories which is not DUSAF or GL30)*

12. A pixel can be played only once by the same player.

5.3 Design and architecture of Land cover validation game

The design of Land Cover Validation is based on the Open Geospatial Consortium’s (OGC)[17] Standards-based geospatial portal. This Geo-Portal Portal Reference Architecture depends upon the principle of service Oriented Architecture (SOA), where services are discoverable on a network, facilitating data integration and interoperability (Erl, 2005). The Geospatial Portal Reference Architecture sketches four classes of service that has to be satisfied in order to be OGC compliant: portal, portrayal, data and Catalogue services. The portal service incorporated in the Land Cover Validation game serve as system management. The portrayal service is implemented in the form of a Web Map Service[18] (WMS), which displays the base layer (i.e.) orthophoto for the validation of pixels. The data service consists

[17] *Open Geospatial Consortium (OGC), an international voluntary consensus standards organization.*

[18] *A Web Map Service (WMS) is a standard protocol for serving georeferenced map images over the internet that are generated by a map server using data from a GIS database. The specification was developed and first published by the Open Geospatial Consortium in 1999.*

of pixel coordinates to create GeoJSON[19] layer on the map. The Catalogue Web Service (CWS)[20] is a metadata portal which gives the information about the data.

5.3.1 General framework and System components

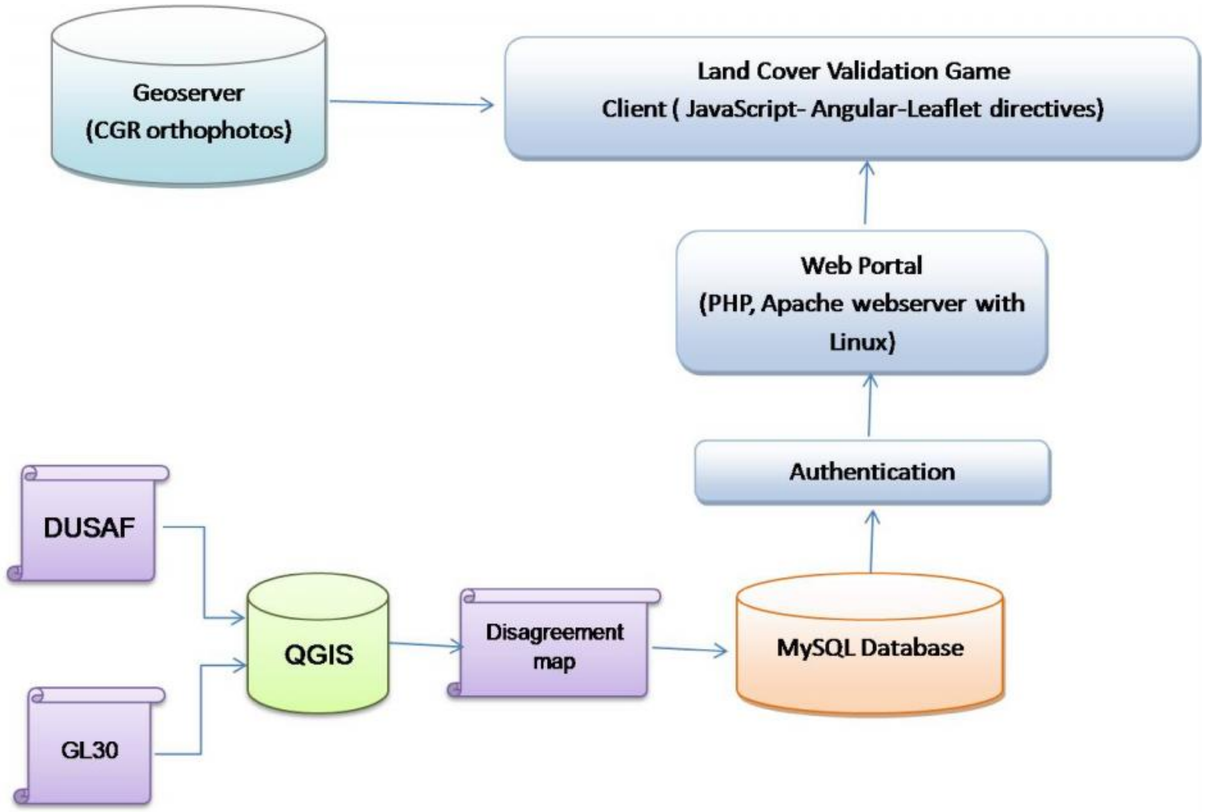


Figure 21: Architecture of the Land Cover Validation Game.

The figure 21 gives the schematic diagram of Land Cover Validation architecture, which consists of many standard components unified into a single portal. The two different repositories are used in Land Cover Validation game to store base map layer and data products, which are described in the section 5.3.5. The first repository (Geoserver) contains the ortho photo which was kindly provided by CGR. S.p.a for the two baseline years 2000 and 2010. These datasets are displayed in the game a Web Map Service (WMS) by making use of the leafletjs[21]. Leaflet is used to build Web mapping application and it is widely in use because of its simplicity. The second repository contains disagreement pixel coordinates and

---

[19] *GeoJSON is an open standard format for encoding collections of simple geographical features along with their non-spatial attributes using JavaScript Object Notation. The features include points, line strings, polygons, and multi-part collections of these types.*

[20] *Catalog Web Service is a standard for exposing a catalogue of geospatial records in XML on the internet (over HTTP). The catalogue is made up of records hat describe geospatial data, geospatial services, and related resources.*

[21] *Leaflet is a widely used open source JavaScript library used to build web mapping applications.*

a database of Land Cover Validation game user, their contribution, the score obtained, badges awarded. More information about the Database is given in section 5.3.2.

The Web portal functions using PHP running on an Apache Web Server (version 2.2.22) together with the WMS. The server runs on Ubuntu machine. The client browser loads the game, which is written in PHP and JavaScript that are purely open source. Each time the player answers one particular pixel, the client computer initiates request for another pixel in JavaScript to Database and then GeoJSON are displayed on the Orthophoto. The high resolution orthophoto helps in better differentiating the classification of land cover.

5.3.2 Database Design

The database design in the Land Cover Validation game is made simple and well-structured using open source MySQL relational database management system. The figure 22 shows the ER model of Database system of the Land Cover Validation game. The database stores the user details, resource (pixel coordinates), DUSAF-GL30 classification, badge lists, user validations, game round detail, Leader board. The core tables which are very vital for this game are Resource, Topics and Resource_has_topic. These three tables influence the major part of the game. The following tables list the information which is stored in it.

1. The *User table* store the first name, last name, type of social network they used for logging, id of their social network and access token.
2. The pixel coordinates are loaded in the *resource table*, including their DUSAF and GL30 classification.
3. The five categories of first level CORINE Nomenclature (see chapter 4) are stored in *Topic table.*
4. A junction table "*Resource_has_Topic*" contains links information between specific pixel and specific categories along with their confidence limit.
5. The *badge table* contains names, description and image of each badge.
6. The *leader board table* stores the cumulative score of each player of Land cover validation game.
7. Once the player starts playing all his chosen categories for each pixel is stored in *logging table*. The Logging table stores the id of the user, the id of the pixel, id of the chosen categories, the confidence level of the particular pixel before his/her play, reputation of the user during the previous round and id of the round he/she plays. This table is very important for future validation processing.
8. The score obtained in each round for each player and the time stamp of the beginning and end of each round is stored in *round table*.
9. The badges assigned during the game play for each player is stored in *user_has_badge table* along with the time stamp.

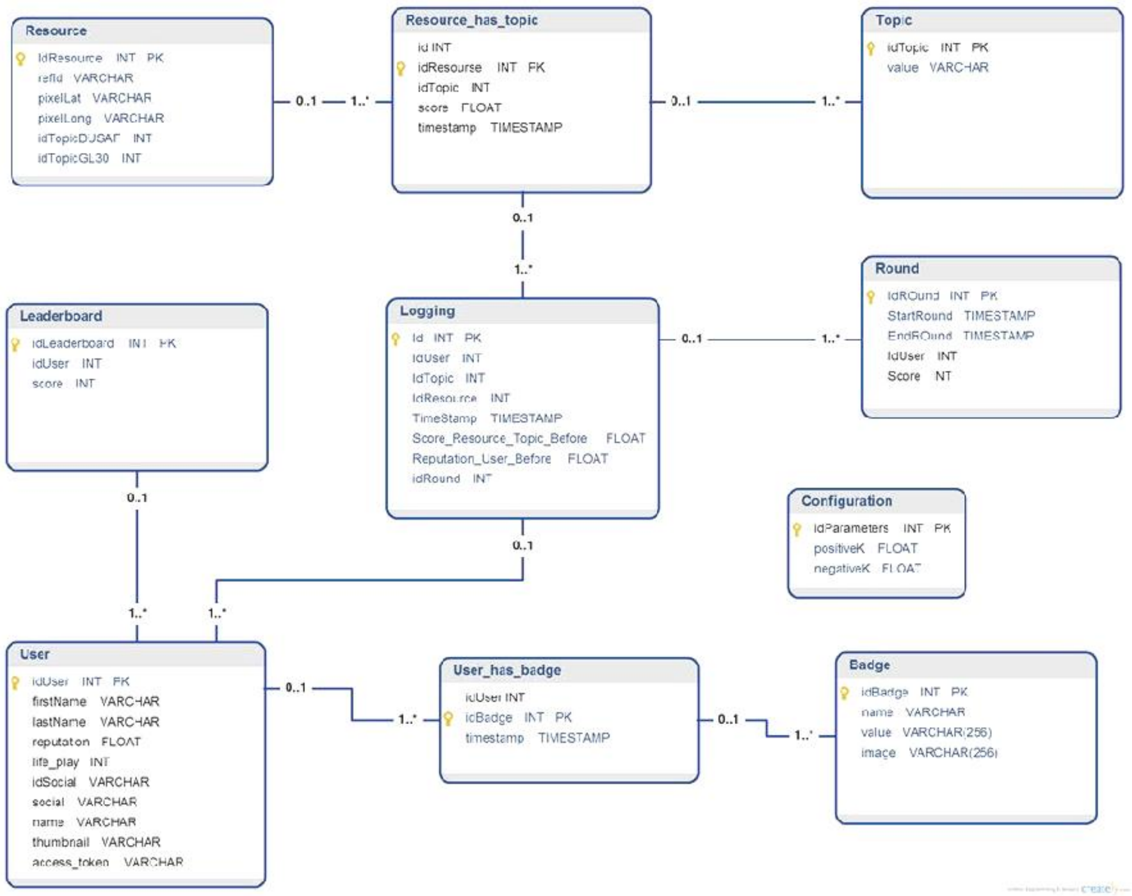


Figure 22: Entity Relation model of database system for the Land Cover Validation game.

5.3.3 Use Case diagram

A Use Case diagram is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved.

5.3.3.1 Player Use Case diagram

The Use Case diagram for the player is shown in the figure 23, which explains how the player can interactive with different module within the game to play and perform validation.

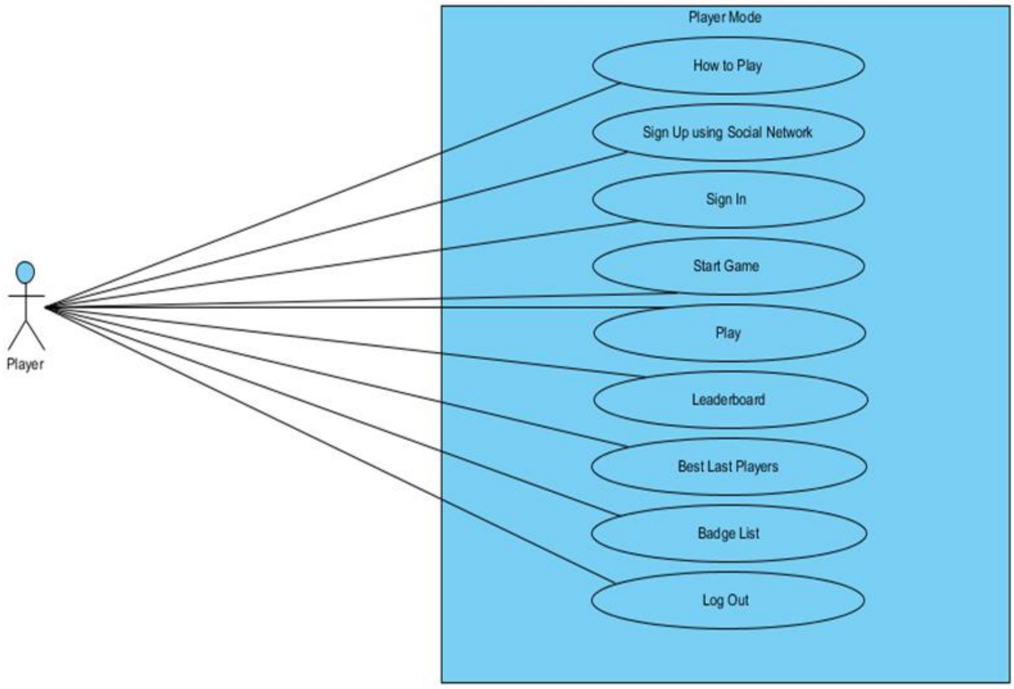


Figure 23: Player Use Case diagram.

5.3.3.2 Admin Use Case diagram

The Use Case diagram for the Admin is shown in the figure 24, which explains how the Admin can interactive with different module within the game to prepare the game to the player and how to obtain the results from the validation.

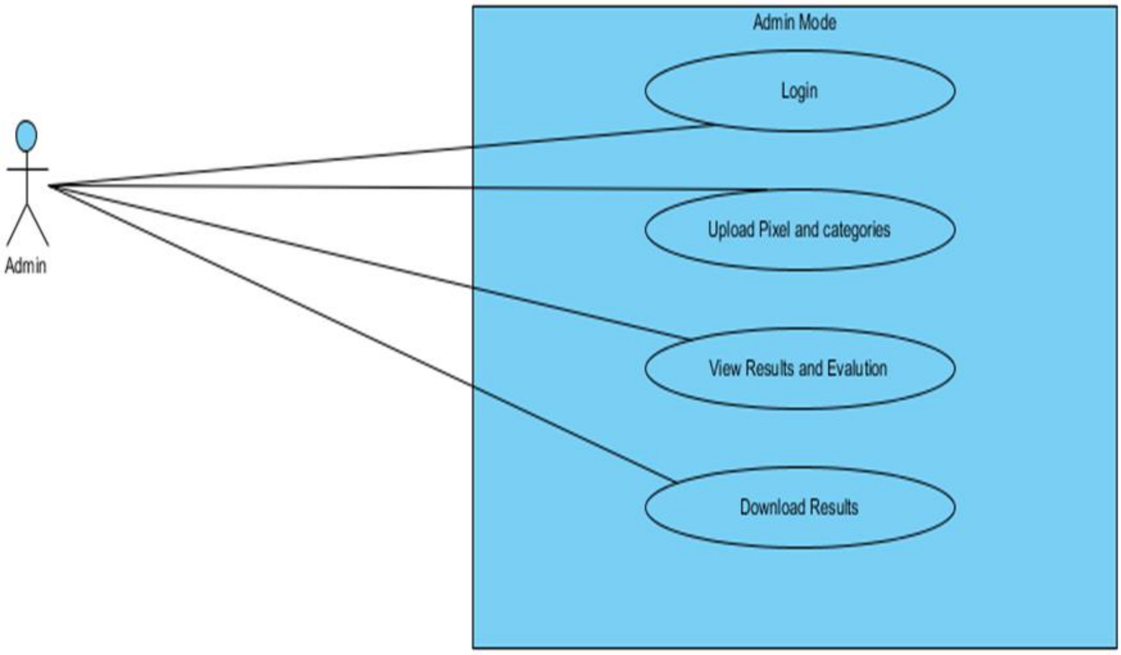


Figure 24: Admin Use Case diagram.

5.3.4 Operation flow

5.3.4.1 Activities diagram

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modelling Language, activity diagrams are intended to model both computational and organizational processes (i.e. workflows). Activity diagrams show the overall flow of control. The figure 25 gives the activity diagram of game play and figure 26 gives the activity diagram of admin operations.

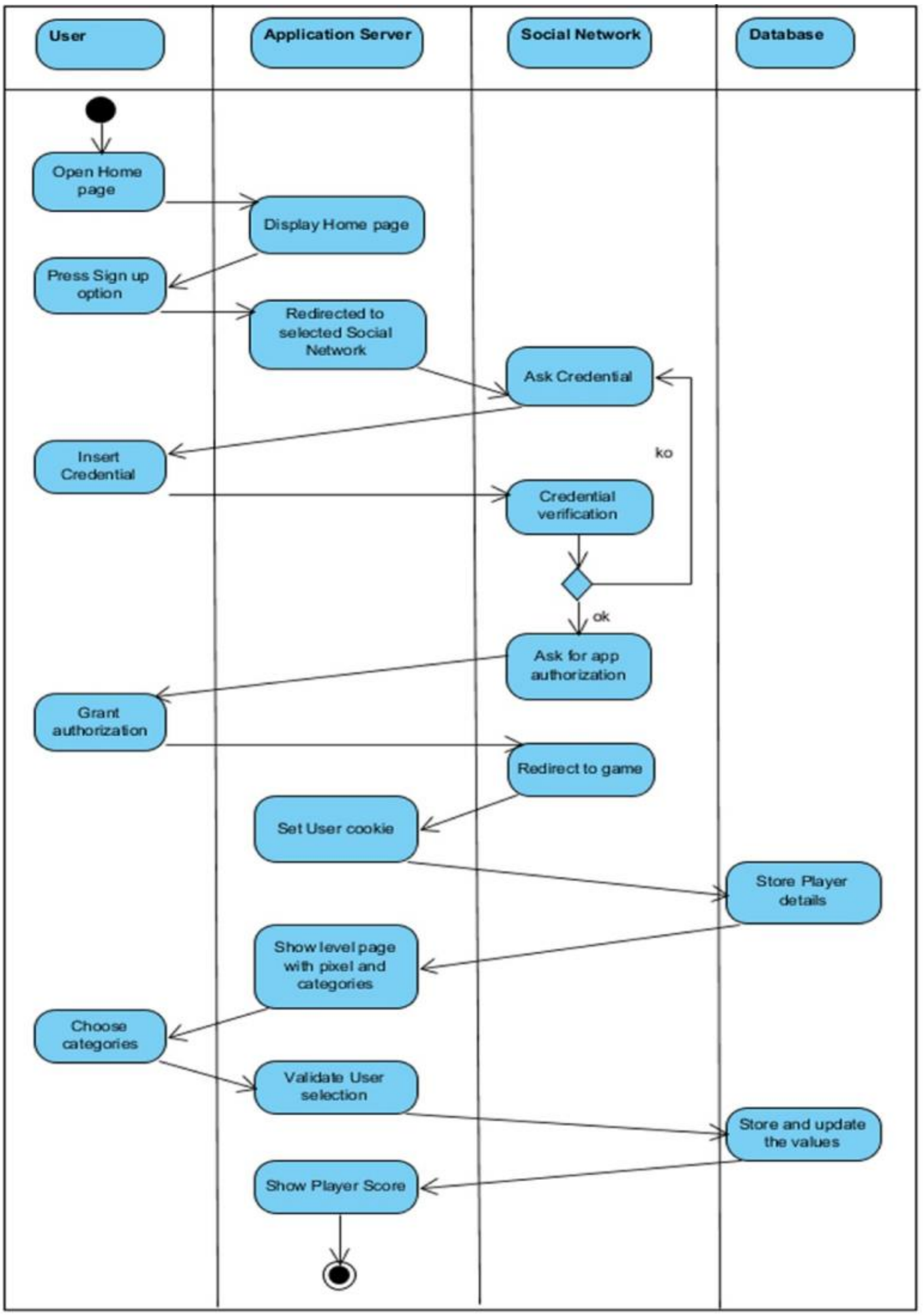


Figure 25: Activity diagram of game play.

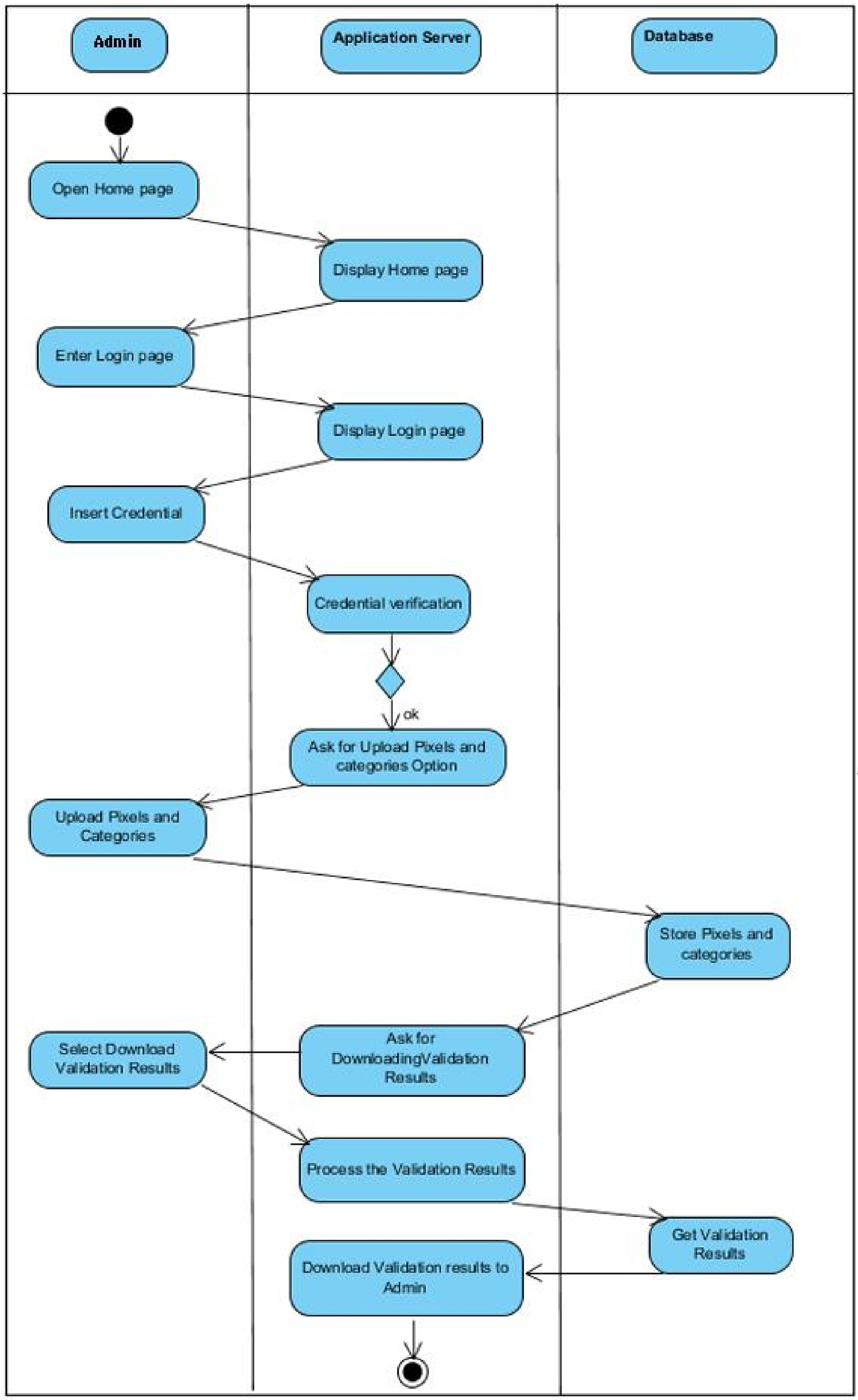


Figure 26: Activity diagram of Admin operation.

5.3.4.2 Sequence diagram

A Sequence diagram is an interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. A sequence diagram shows object interactions arranged in time sequence. The figure 27 shows the sequence diagram of game play. The figure 28 shows the sequence diagram of admin operation.

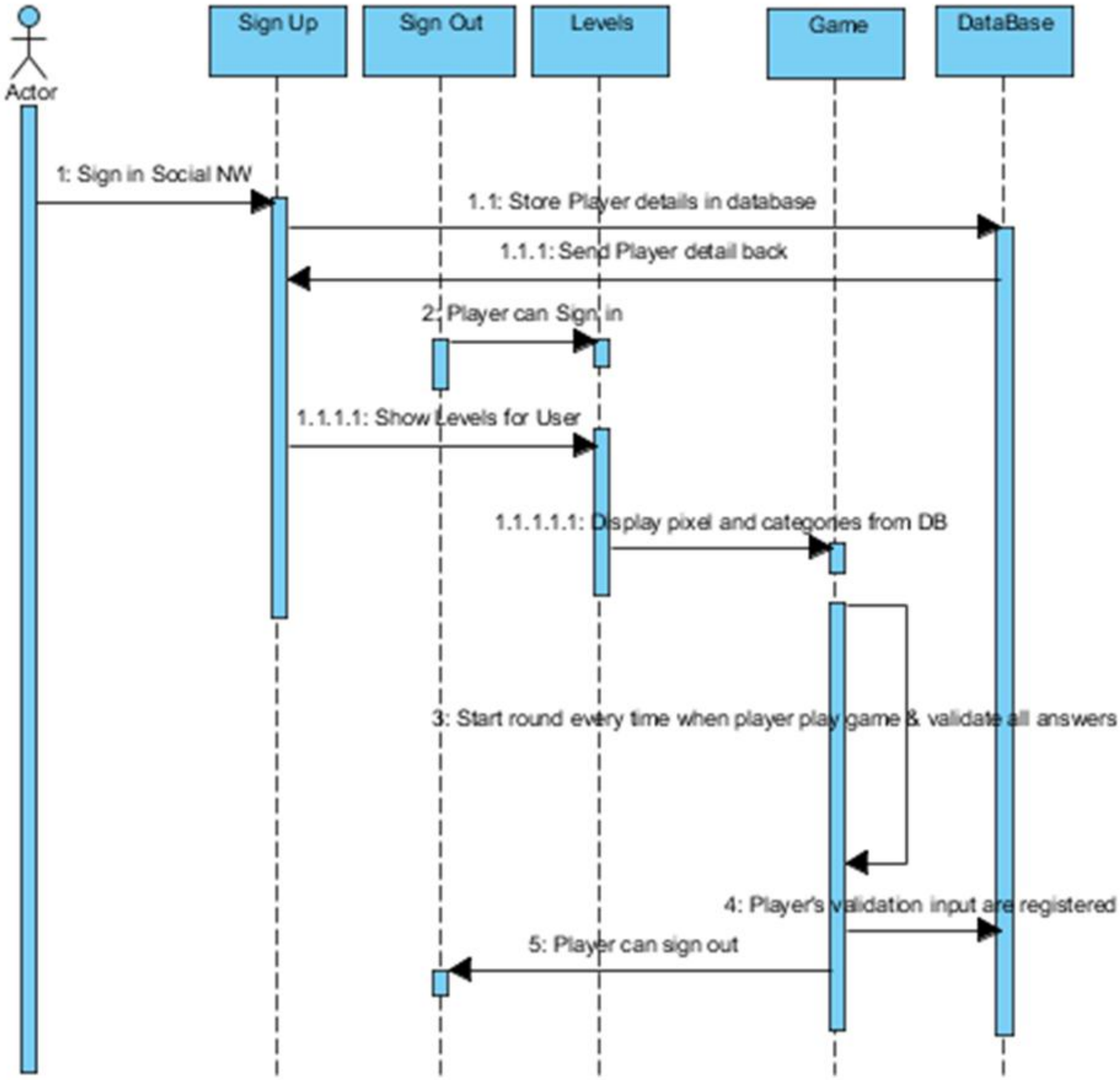


Figure 27: Sequence diagram of Game play.

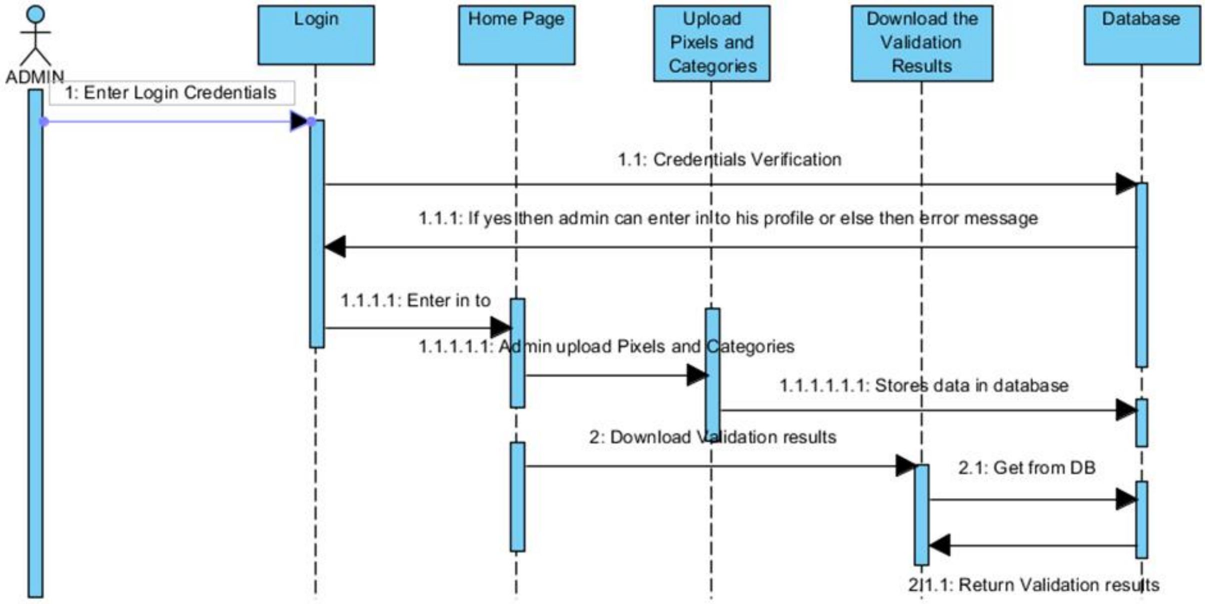


Figure 28: Sequence diagram of Admin.

5.3.5 Base layers and data products

Land Cover Validation game displays pixel above the base layer (i.e.) orthophoto of 1m resolution kindly provided by CGR s.p.a[22]. These Orthophoto tiles are made available in the game using the GeoServer. During the testing phase, we found that Geoserver was crashing very frequently due to the large request from the many players during the validation process. In order to fix the problem we tried to look into the Geospatial Data Abstraction Library (GDAL)[23] information of the file using the functional command *gdalinfo filename.tif.* Some of the problem sorted out by viewing the GDAL information such as

1) Missing Coordinate Reference System (CRS).

2) Missing Geo-referencing.

3) Bad Tiling.

4) Missing Overviews.

5) Compression.

The detailed version of code to pre-process the orthophoto has been given in the Appendix A. The following step briefs the pre-processing procedure.

---

[22] *http://www.cgrspa.com/*

[23] *The GDAL is a library for reading and writing raster geospatial data formats, and is released under the permissive X/MIT style free software license by the Open Source Geospatial Foundation.(http://www.gdal.org/)*

STEP 1: Fix and Optimize *gdal_translate*[24]

STEP 2: Add Overview with *gdal_addo*[25].

The disagreement pixel has been obtained as shown in the section 4.3. These pixel coordinates have been stored in MySQL database. Using these pixel coordinates, the GeoJSON[26] layer is created and displayed over the orthophoto.

5.3.6 Social Networking Authentication

This topic describes how the user credentials validations done through social network service Application Programming Interface (API)[27] (Facebook, Twitter and Google-Plus). We make use of HelloJS[28] which Standardized paths and responses to common API's including Facebook, Twitter and Google Plus.

5.3.6.1 Creation of Facebook Application ID

The following processes are the steps to create Facebook Application ID.

1. Go to https://developers.facebook.com/ page and "add new app" in My Apps tab, then select the platform.

2. Give new app name "Land Cover Validation game" and create new Facebook app id then choose category as Education and click on "Create app id".

3. Now URL of the game is added.

4. Select the "Land Cover Validation game" under My Apps.

5. Now App id is created which is used inside the application for the authentication.

6. In order to make the app live and available to the general public in status & Review of the app choose Yes to make the app available to general public.

[24] *The gdal_translate utility can be used in converting the raster data between different formats, and can be used in performing different operations including subsettings, resampling and rescaling pixels in the process*

[25] *The gdaladdo utility can be helpful in building or rebuilding overview images for the most supported file formats with different downsampling algorithms*

[26] *GeoJSON is an open standard format for encoding collections of simple geographical features including non-spatial attributes using JavaScript Object Notation.*

[27] *Application Programming Interface is a set of routines, protocols, and tools for building software applications.*

[28] *http://adodson.com/hello.js/*

5.3.6.2 Creation of Twitter APP ID

Creating App Id in Twitter is similar to creating App Id in facebook.

1. Go to https://apps.twitter.com/ page and create new app.
2. Give the required information and website and create app id.
3. The additional part in twitter is to configure the app id in HelloJS. This can be managing apps in https://auth-server.herokuapp.com/#-auth-server
4. Add the app credentials including client_id, domain, client_secret from the twitter

5.3.6.3 Creation of Google Application ID

The following processes are the steps to create Google Application ID.

1. Go to Google Developers Console https://console.developers.google.com/ and create new project.

2. In APIs & auth → APIs add Google+ API

3. In Credentials tab, Create new Client ID, enter game link and callback URL.

4. In Consent screen, enter the product name and email id.

5. The created client ID is being used in the application to authenticate the user.

After extracting the App ID from these three social networks, it has been added in the index.html page.

5.4 Game - Web Application development

5.4.1 Front-end development

The User-Interface of Land Cover Validation game is developed based on linked data technologies[29] using AngularJS[30], an open source web application framework maintained by Google. The AngularJS library works by first reading the HTML[31] page, which has embedded into it additional custom tag attributes. The library defines those tag attributes as directives to bind input or output parts of the page to model which represents standard JavaScript variable.

[29] *In computing, linked data describe a method of publishing structured data so that it can be interlinked and become more useful through semantic queries.*

[30] *https://angularjs.org/ Angular JS is an open-source web application framework maintained by Google and by a community of individual developers and corporations to address many of the challenges encountered in developing single-page applications.*

[31] *Hypertext Markup Language, commonly referred to as HTML, is the standard markup language used to create web pages.*

The framework contents are designed as HTML pages with CSS and for the data transmission JavaScript libraries have been used. There are ten HTML pages, which are designed in which nine pages are visible to the player and one page (Evaluation.html) is hidden to the players. The player can navigate from one page to another page.

The list of .html pages as follows

1. home.html
2. about.html
3. login.html
4. formUser.html
5. level.html
6. complete.html
7. badgeList.html
8. leaderboard.html
9. topTenRound.html
10. evaluation.html (hidden)

The home page (home.html) of Land Cover validation looks as shown in the figure 29. The home page always opens with popup “How to Play” (Figure 30) which brief the player with a list of information as follows:

1. What is given in the game?
2. How to choose the categories by looking at the pixel.
3. How to manage with the challenging pixels (i.e.) mixed pixels.
4. How the score is assigned to the players.

The “Start New Game” button is placed in the left panel which is not dynamic.

Figure 29: The home page of Land Cover Validation game.

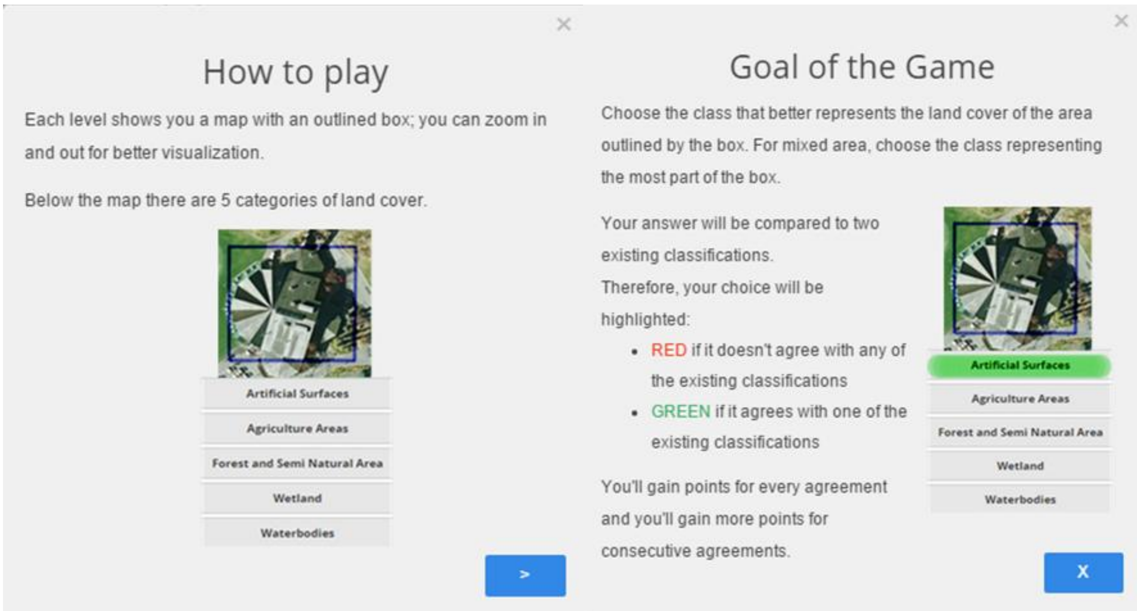


Figure 30: Game information Popup.

The player can navigate between different pages. "What is this about" (about.html) page (Figure 31) describes the following things as follows.

1. What is main objective of this game?
2. What are the different data products used?
3. What methodology is adopted for the validation?
4. Which base map is utilized for the purpose of validation?
5. This page also provides the user with the links to download the data used in the game which is available for free.

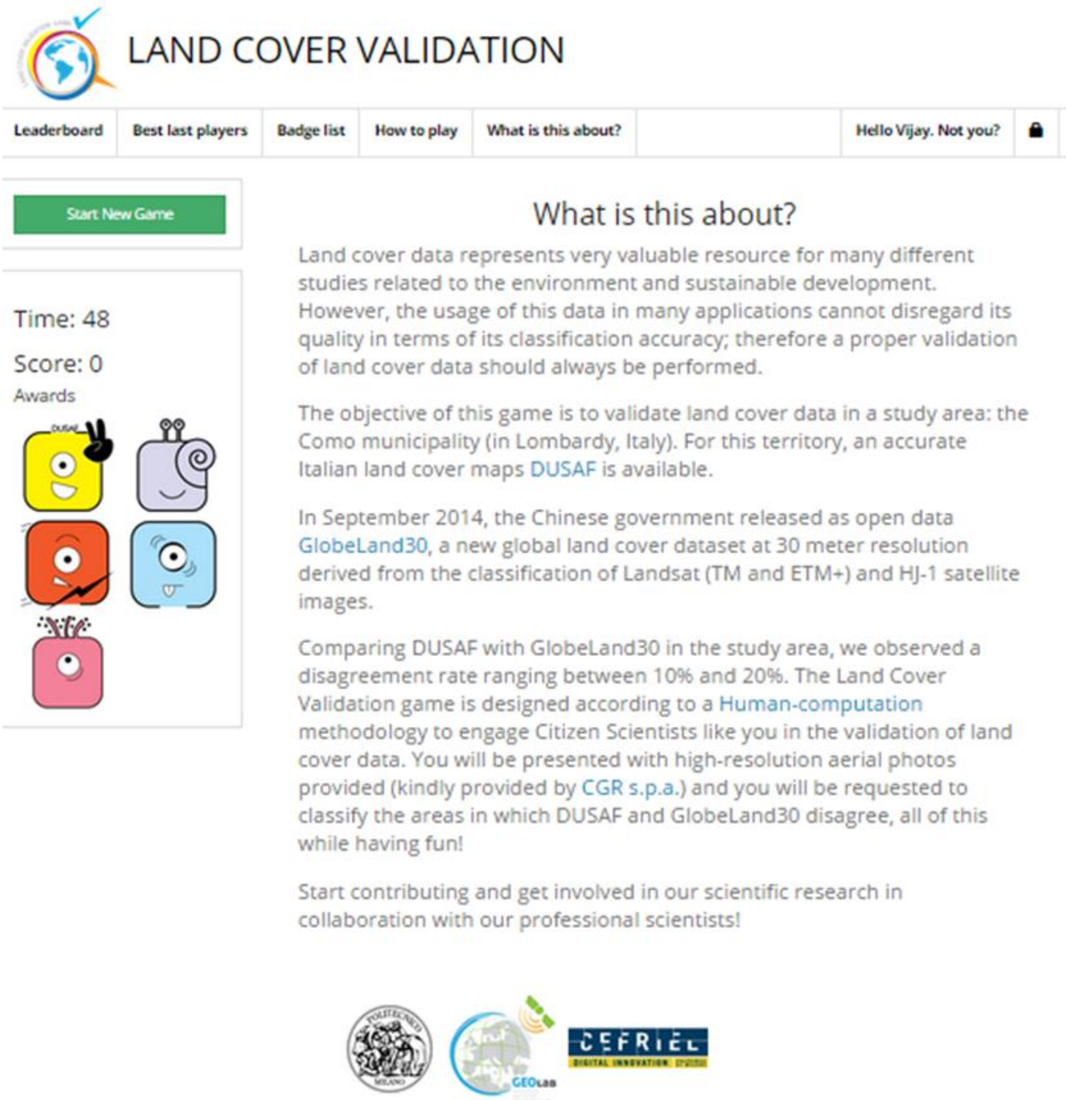


Figure 31: The about page of Land Cover Validation game.

The next page is the “User form” (form.html) page (Figure 32), when a player hits “Start New Game” the home page is directed to “User form” page which contains three Social Networking service buttons to be starting the game.

Figure 32: The User form page.

When a player completes single round of the game he will be displayed with the complete (complete.html) page (Figure 33) where he/she will be displayed with the score, assigned badges (if any), how many points needed to be first in the Best Last Players and again start button to continue the game to the next round.

In the left panel time, score and badges assigned for the current player are displayed.

Figure 33: Complete page.

The main game page (Figure 34) can be seen using *level.html* which is the dynamic page, the contents of the page changes dynamically when player elevate to higher levels.

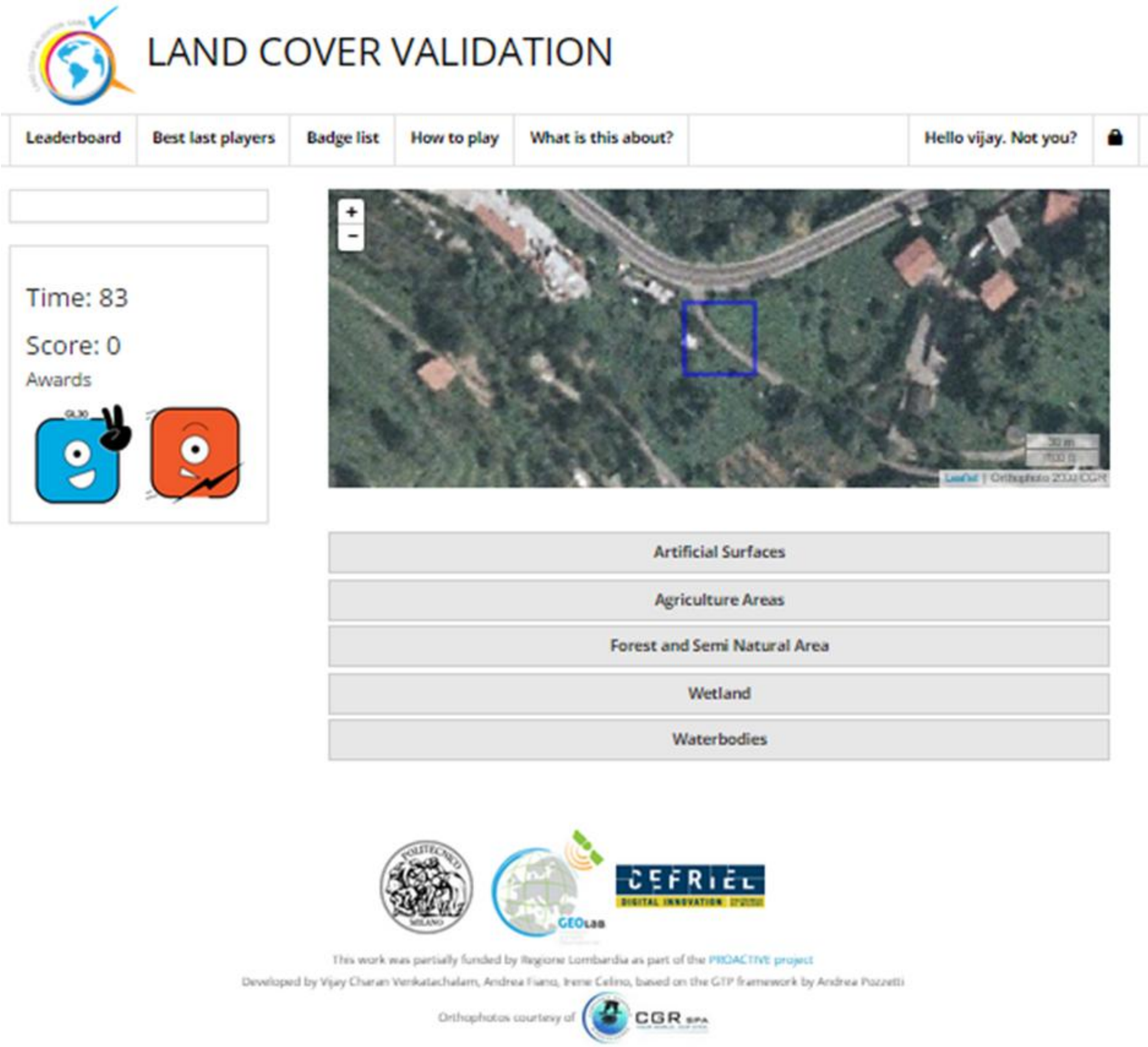


Figure 34: Level page.

The leader board (leaderboard.html) page (Figure 35) list the players according to the cumulative score which they obtained during the game play. It displays top three people who are leading in the game as well as the player position who is playing.

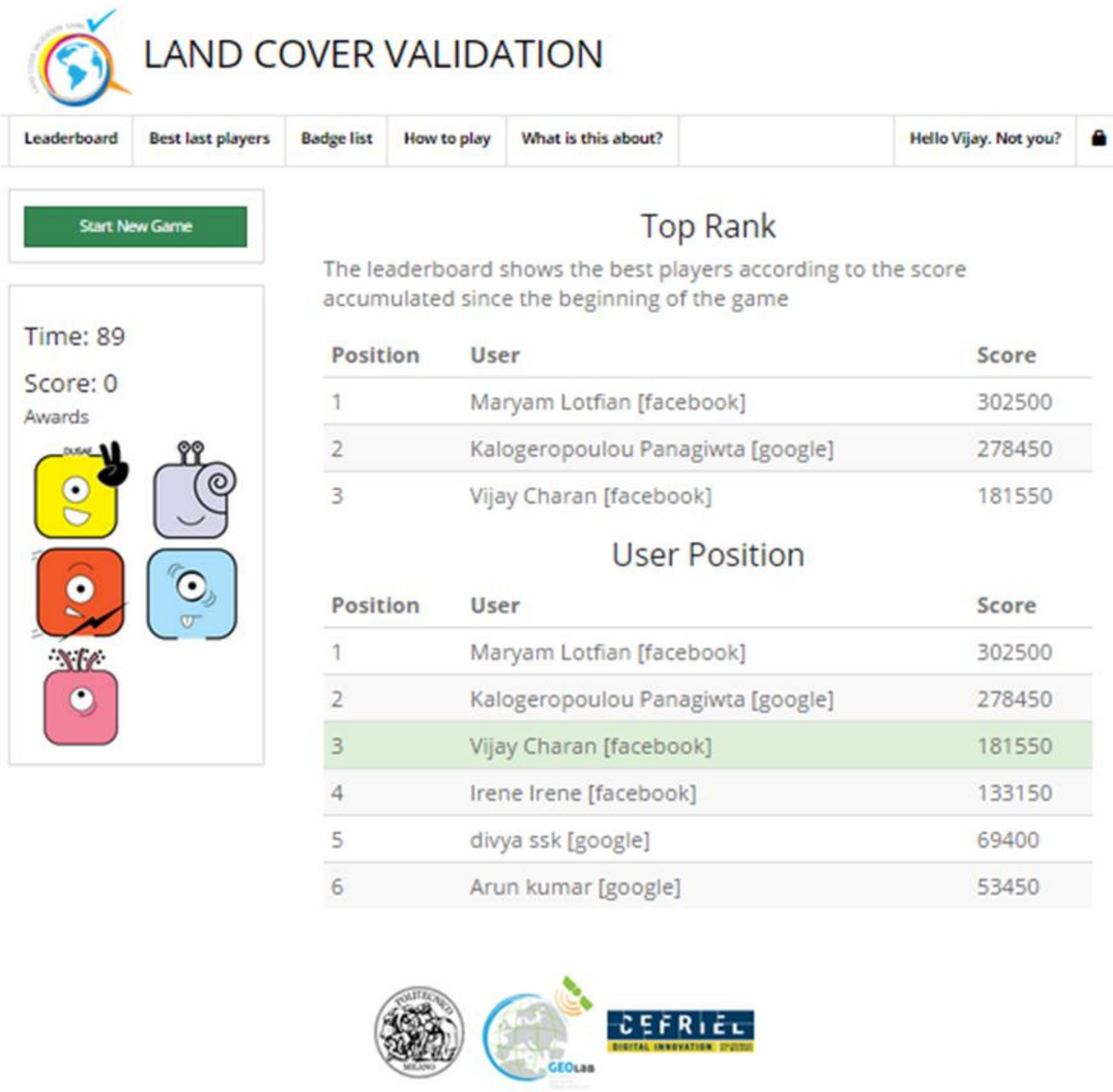
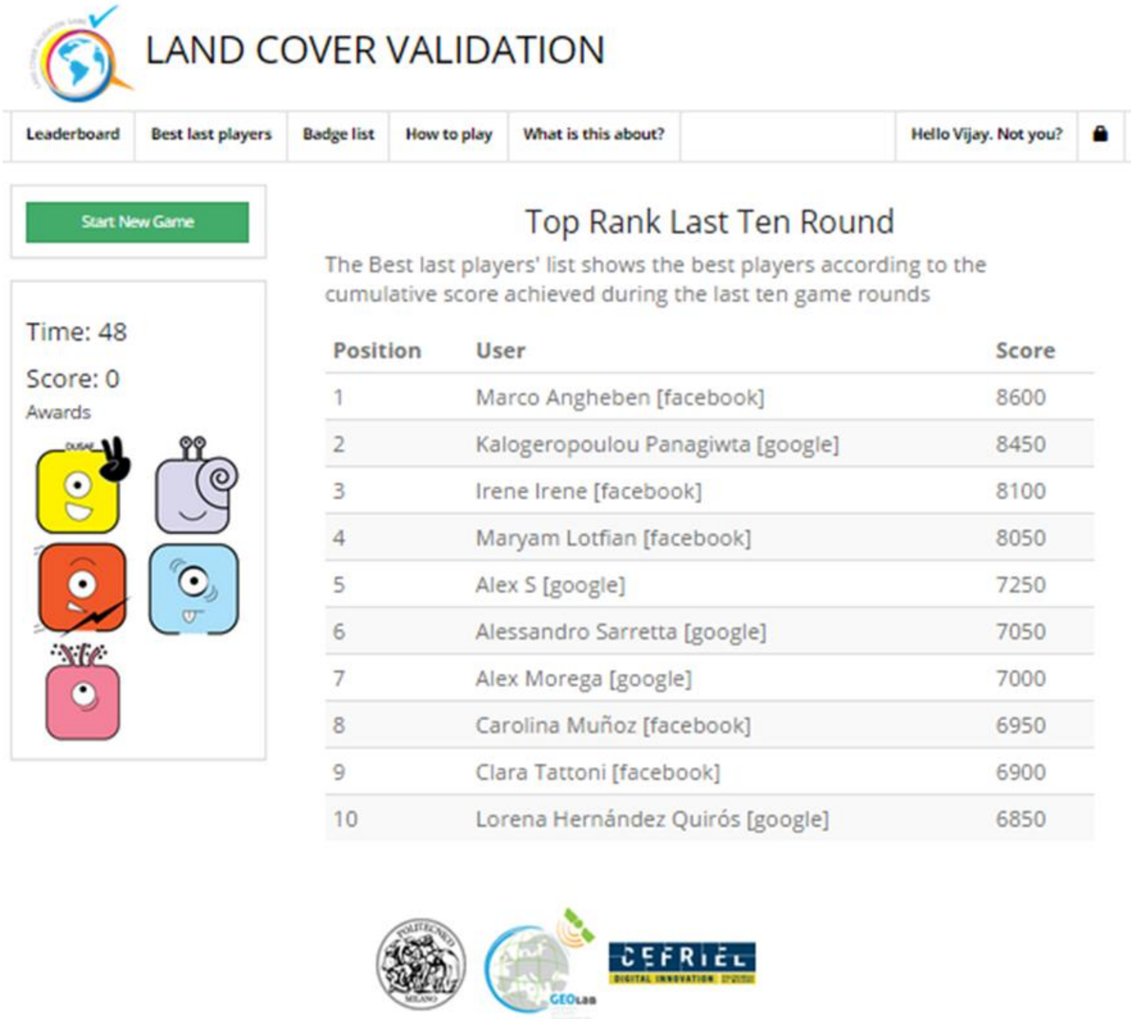


Figure 35: The Leader board page.

The Best last players (topTenRound.html) page (Figure 36) displays the position of the players according to the cumulative score obtained during the last ten game rounds.



Position	User	Score
1	Marco Angheben [facebook]	8600
2	Kalogeropoulou Panagiwta [google]	8450
3	Irene Irene [facebook]	8100
4	Maryam Lotfian [facebook]	8050
5	Alex S [google]	7250
6	Alessandro Sarretta [google]	7050
7	Alex Morega [google]	7000
8	Carolina Muñoz [facebook]	6950
9	Clara Tattoni [facebook]	6900
10	Lorena Hernández Quirós [google]	6850

Figure 36: The Best last players page.

The badge list (badgeList.html) page (Figure 37) displays the list of badge lists and the description of each badge.

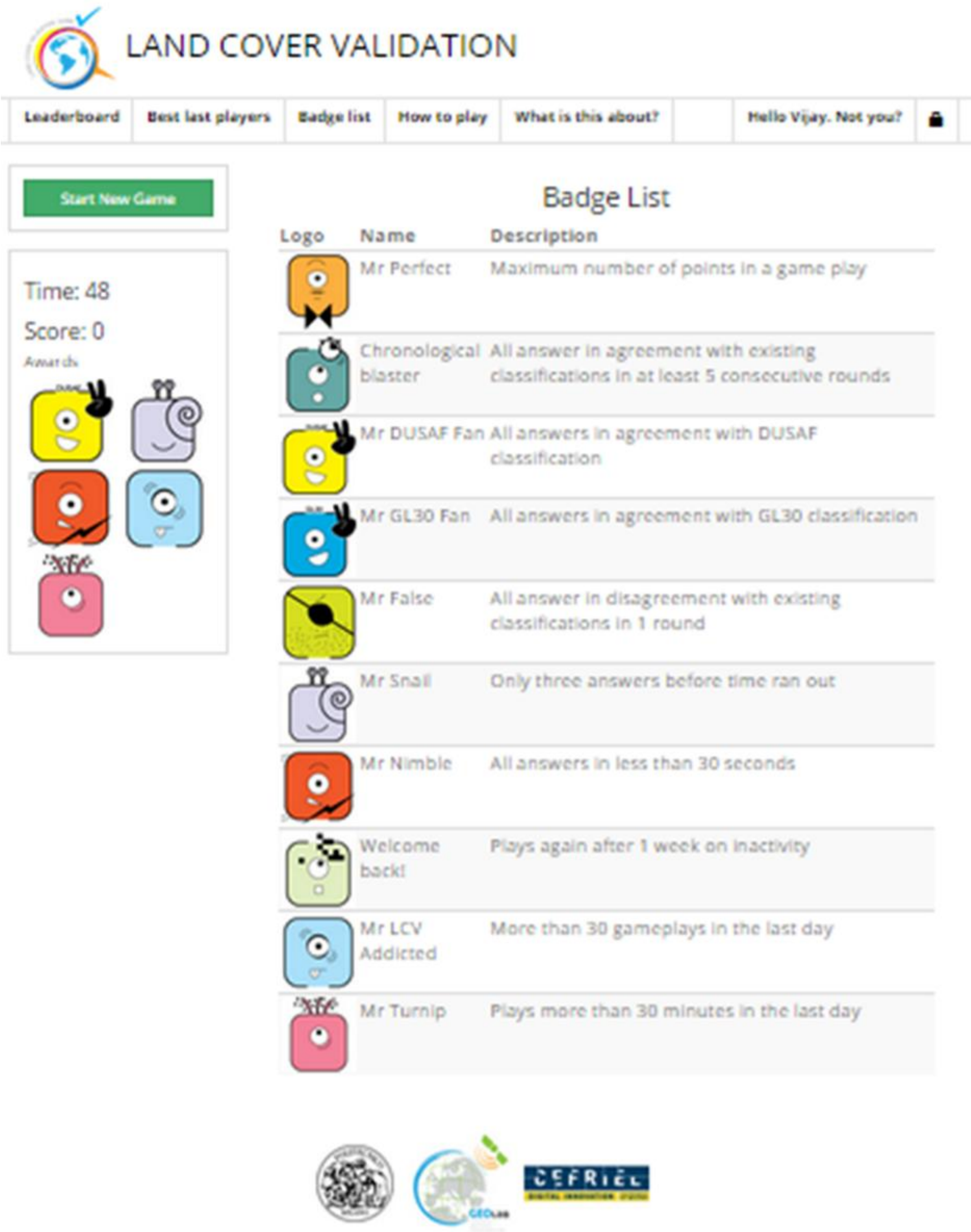


Figure 37: The Badge list page.

In order to make client-side scripts to interact with the users in web to control the browser, communicate asynchronously, and alter the document content that is displayed we make use of JavaScript, which is a dynamic programming language, used as part of web browsers. JavaScript is also considered an "assembly language of the web". Here we used different JavaScript libraries to display base layers, social authentication purpose and front end development. Some of the important functionalities have been listed here.

To fetch the Orthophoto from the Geoserver and display it in the game, we used Angular-Leaflet directives which perform the above mentioned process. The function to load the map in the html page is given in the Appendix B.

Then for each level new pixel is loaded (i.e.) the map is panned to another location according the pixel coordinates (see Appendix B) which is obtained from the Database. The pixel is marked in blue coloured GeoJSON layer.

5.4.2 Back-end development

The back-end of the Land Cover Validation game is 16 different APIs. These 16 APIs drives the application to perform various computations within the game. Those are listed below as follows

1. Admin.php
2. badgeList.php
3. complete.php
4. db_connect.php
5. evaluation.php
6. exportCSV.php
7. functions.php
8. leaderboard.php
9. levels.php
10. logging.php
11. setParameters.php
12. topTenRound.php
13. Upload.php
14. userBadgeIdRound.php
15. userID.php

16. username.php

Of these 16, complete.php and function.php plays the major role in the computation of the game. The Admin.php performs the operations including insertion of pixels for the game play and to alter initial threshold. The badgeList.php connects the badge list table from the MySQL DB to badge list html page. The complete.php evaluates the player's score from the MySQL DB and makes it available in complete.html page. The db_connect.php helps to connect the land cover DB to the application. The evaluation.php evaluates the percentage of agreement pixels with DUSAF, GL30 and disagreements from the MySQL DB and makes it available in evaluation.html page. The exportCSV.php helps the admin to add pixels in game and store them in DB. The different functionalities which are down using PHP are explained in detail in Appendix B.

5.5 Implementation of Land Cover Validation Game

5.5.1 Validation Process

Each pixel in the land cover validation game is validated by the Citizen Scientist based on increasing its confidence level. When a particular pixel reaches the upper threshold confidence limit after getting validated from a series of reputed player, it is no more available in the game for the other players.

5.5.1.1 Initial Confidence level

The initial Confidence level of each category (topic) is pre-defined in the game (See figure 5-18). The confidence level of DUSAF is set as 65% (0.65 in the game); we came up with this value on the basis of some domain-dependent[32] condition and design choice (Celino et al. 2012)[lxiii]. We believe DUSAF is more reliable than GL30 since it has been produced using high resolution aerial photographs. The initial confidence level of GL30 is set at 35% (0.35 in the game) which is less when compared with DUSAF, since this land cover maps are produced from Landsat (TM and ETM+) and HJ-1 satellite images which has the lesser resolution when compared with aerial photographs. We followed the same rules as mentioned for DUSAF weighted condition initialization. The initial confidence level of categories which are neither DUSAF nor GL30 is set as zero.

5.5.1.2 Threshold level

The upper and lower threshold level is set for the validation of pixels (Figure 38). This threshold limit is pre-defined in the game in order to eliminate the pixels from the game, which are played more times by the user which are considered to be validated. The upper threshold level is set at 94% (0.94 in the game). The lower threshold is set to 0.04. This kind

[32] The Domain of dependence is a portion of the range such that the initial values on this portion determine the solution over the entire range.

of threshold or weighted voting scheme is set up for aggregating human outputs; this is a usual approach for this kind of cooperative systems[33] (Edith Law and Luis von Ahn 2011)[lxiv].

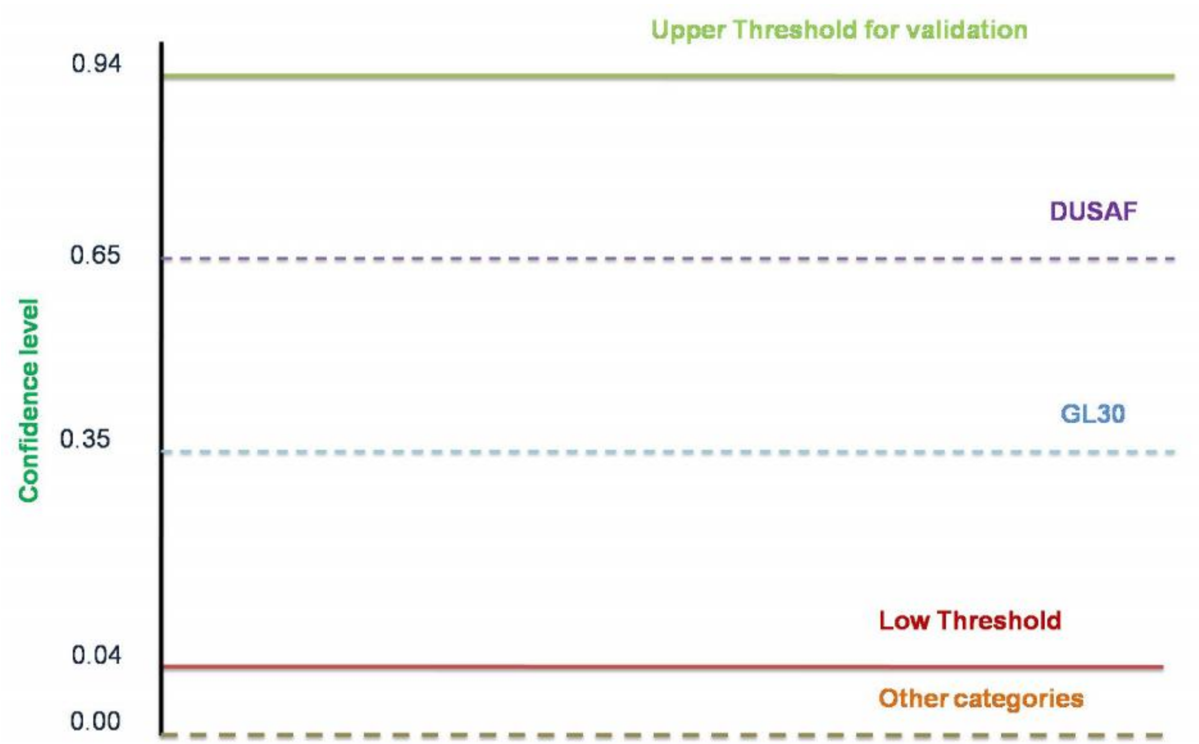


Figure 38: Confidence level of each categories and validation threshold.

5.5.1.3 Methodology

If a player chooses one category (topic) between DUSAF and GL30, the confidence level (Equation 15) of the chosen category increases. Simultaneously the confidence level of the other one decreases. The confidence level alteration is given theoretically as follows

Confidence level of a category = Previous Confidence level + Player reputation * Kpos|neg

Equation 15: Confidence level operation equation for the DUSAF and GL30 category.

Where $|Kpos|=|Kneg|=0.15$

|Kpos| and |Kneg| parameters define the degree of validation and its values are defined based on the rule that the different participants' answers must be aggregated and combined (Law and Ahn 2011) (Celino et al. 2012). Also we wanted at least two participants with reputation 1 to choose the same category to validate them. This 0.15 value can be altered according the developer needs. The greater the value quicker the pixel gets validated. |Kpos| is used to increase the confidence level of the chosen category (between DUSAF and GL30)

[33] *A cooperative system is defined to be a system of multiple dynamic entities that share information or tasks to accomplish a common, though perhaps not singular, objective.*

whereas |Kneg| parameter is used to decrease the confidence level of the category which is not chosen by the player (between DUSAF and GL30).

If a player chooses one of the other three categories which is neither DUSAF nor GL30, the score of the chosen category increases theoretical as shown in the equation 16.

Confidence level of a category = Previous Confidence level + Player reputation * Kpos

Equation 16: Confidence level operation equation for other category.

While the confidence level of DUSAF and GL30 classification remains unchanged.

The important parameter involved in calculation of confidence level is player reputation. The player reputation is computed using the formula as follows

Player Reputation = exp (- Number of errors/2)

Equation 17: Player reputation computation equation.

These formulas are the functional design of GWAP. The player with the higher reputation influence more the confidence level of a category. Practically player with higher reputation can move a category of the pixel to upper threshold easily so that it gets validated.

For a particular pixel when a player chooses the one category which is neither DUSAF nor GL30, the player reputation goes down, but still the confidence level of that particular category increases without affecting the confidence level of other categories. Many players have to agree with the same category for such pixel to move the category to the upper threshold.

The figure 39 shows all the player agrees with only DUSAF category for a particular pixel, so the confidence level keep increasing and reaches the upper threshold level, simultaneously the confidence level of the GL30 decrease and reaches the lower threshold. The figure 40 shows how the confidence level increases for GL30 by decrementing the DUSAF confidence when the entire player agrees with GL30 for a particular pixel.

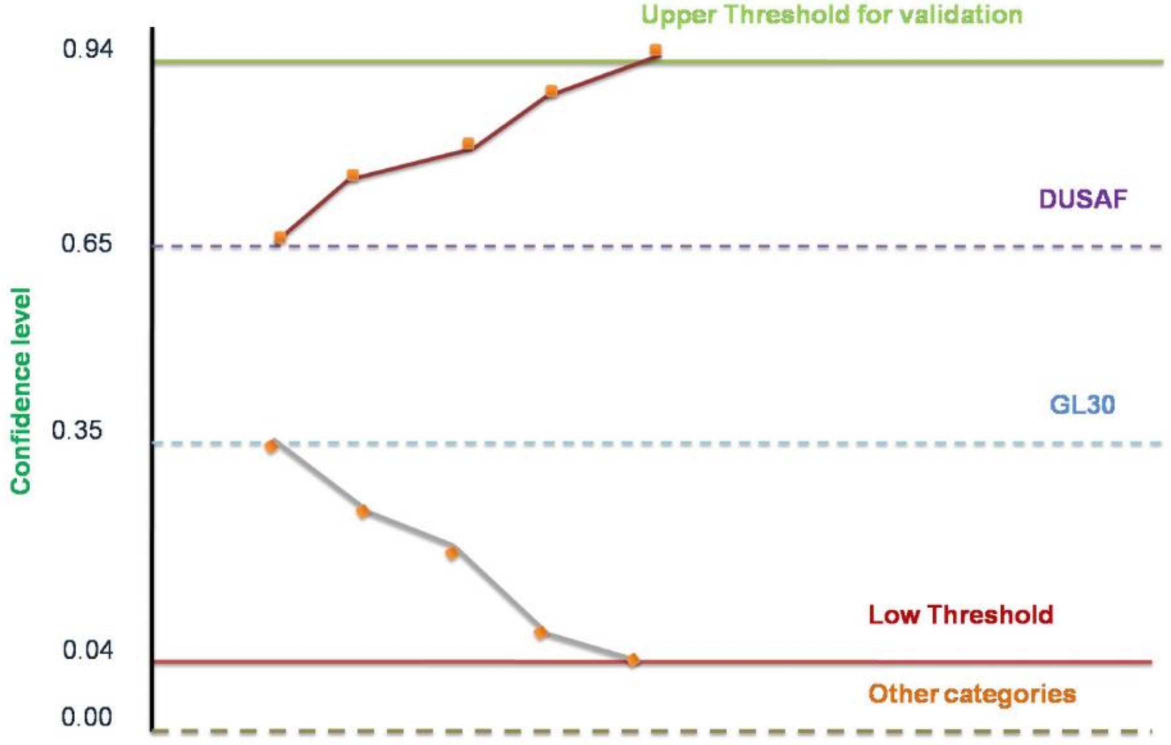


Figure 39: Confidence level increment of DUSAF and decrement of GL30.

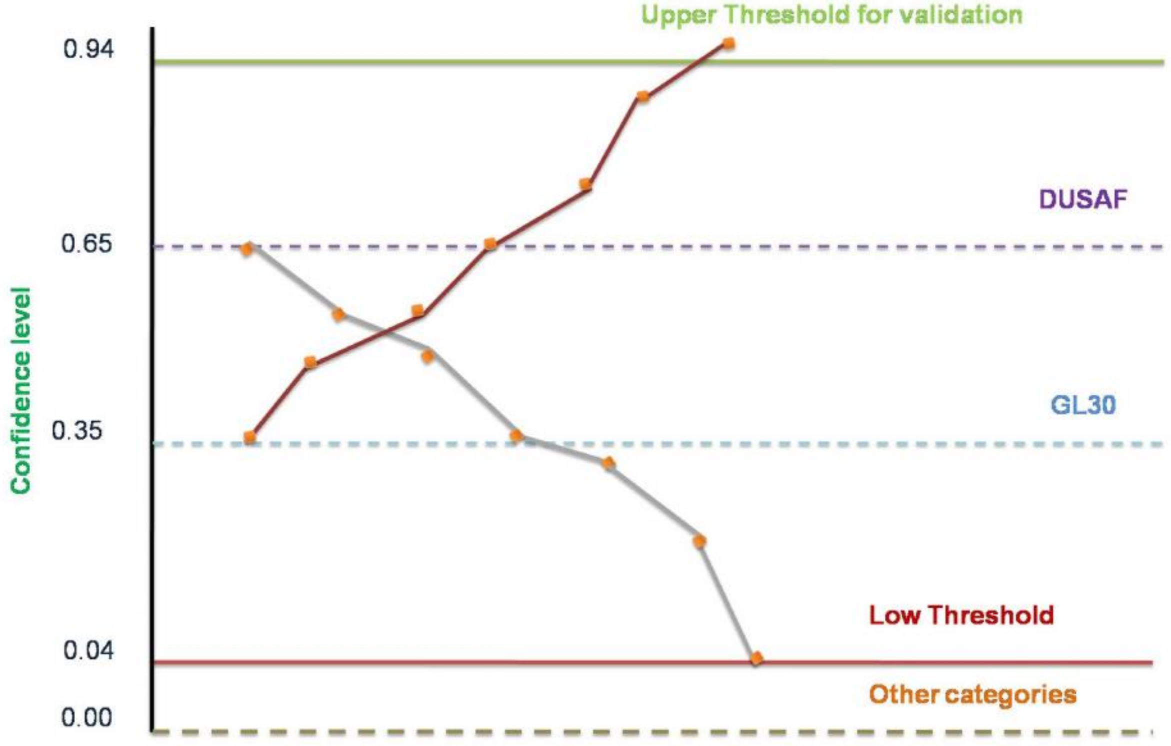


Figure 40: Confidence level increment of GL30 and decrement of DUSAF.

The figure 41shows how the confidence level is contradicting between DUSAF and GL30 classification because of difference between the players opinion for a particular pixel. Here it is clear that the first play increases the DUSAF confidence level thereby decreasing the GL30. During the third game play, the player chosen GL30 hence the confidence level of GL30 started rising by decreasing the DUSAF confidence level. But at the end DUSAF confidence level reaches the upper threshold.

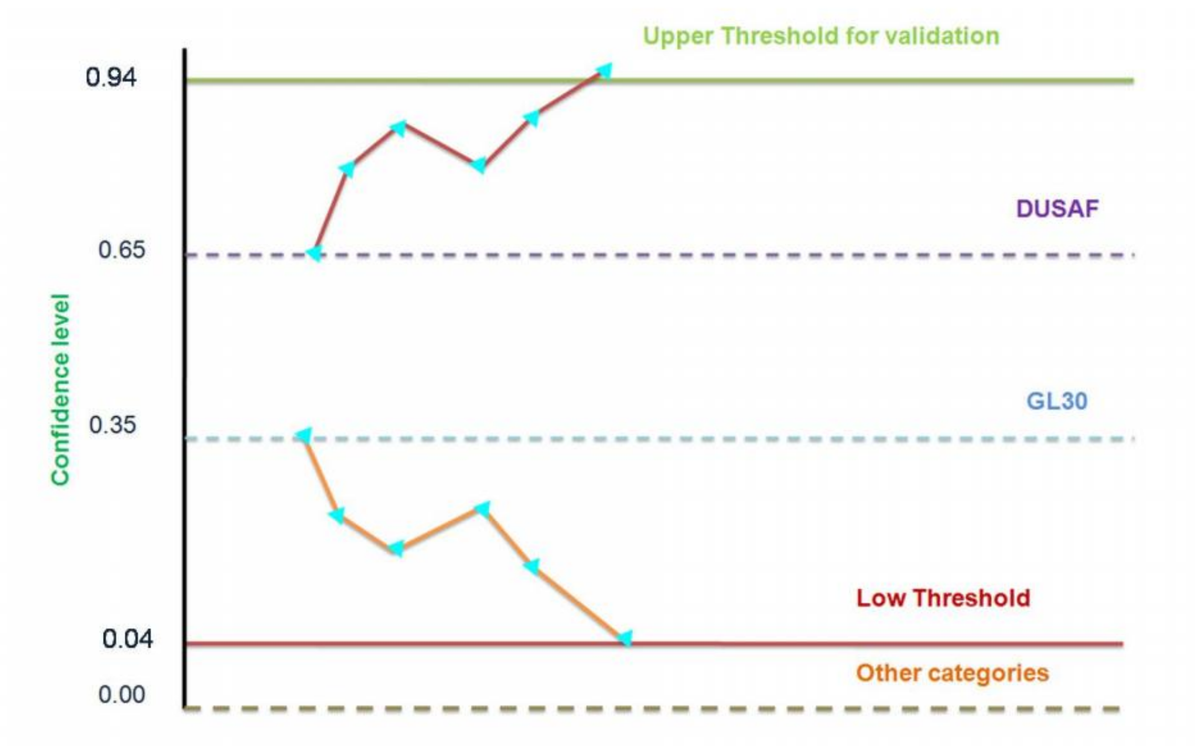


Figure 41: Contradicting confidence level between DUSAF and GL30.

The figure 42 shows how the confidence level of other categories (which is neither DUSAF nor GL30) increases without affecting the DUSAF and GL30 confidence level when maximum player votes for the category apart from the expert answer.

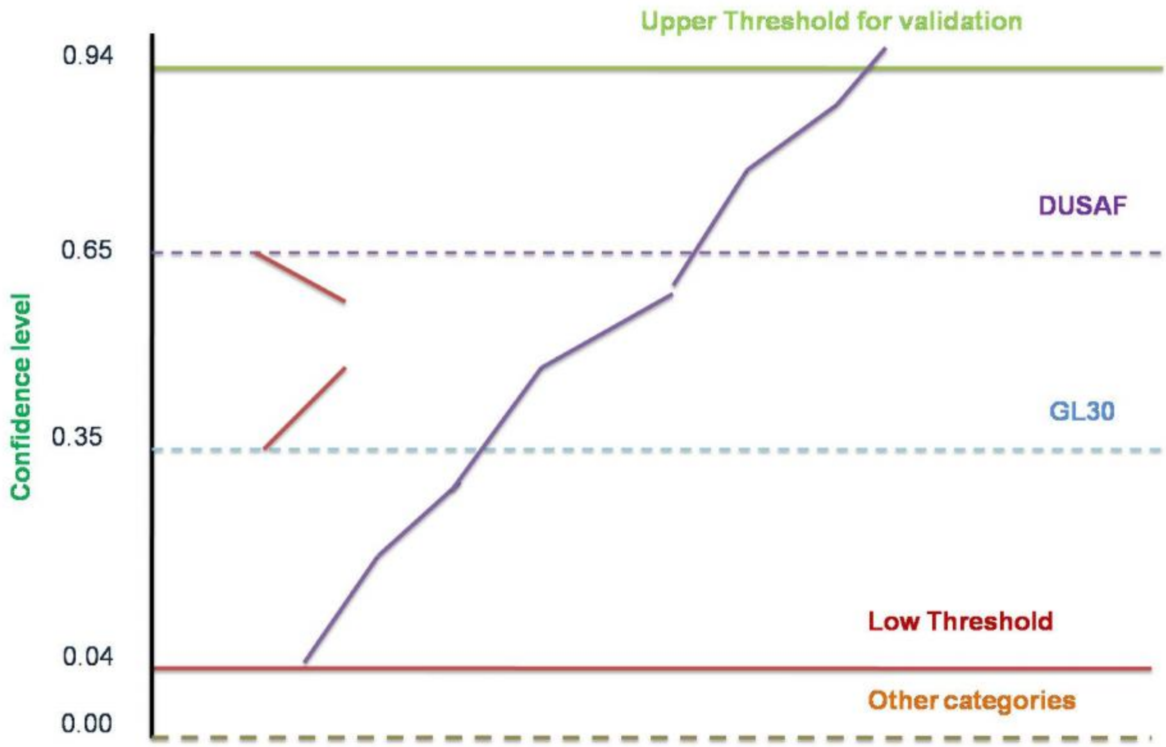


Figure 42: Confidence level increment of other categories.

Chapter 6

6. Testing during mapping party

6.1 Introduction

Mapping party was a part FOSS4G 2015 Europe conference, which took place in Politecnico Di Milano, Como campus. The total participants of the conference were 600. The Land Cover Validation game was one among the four mapping parties of FOSS4G conference. Land cover validation game was introduced on the first day of the conference (15th July) to the participants. It was also published on the FOSS4G website along with the link to play the game (Figure 43).At the end of the conference (17th July) special gifts have been distributed to the top three winners of the game.

At the end of the conference 75% of the game was completed, so the land cover validation was made available online for some more days for the players to play till the game reaches 100% of the competition. The game has been completed fully on 28th July with 1601 pixel fully validated by different users.

Figure 43: Land Cover Validate game published in FOSS4G website.

6.2 Mapping party

Totally 68 players contributed a total of 10438 classification answers, whereas 2032 pixels has been uniquely played since many players can play on the same the pixel till it gets reaches the upper threshold

6.2.1 Initial Result

The directly will be obtained from the game using the evaluation page (Figure 44) which gives the total number of players played, total played time, number of validated pixels, number of played pixels, total number of pixels, Completion rate, Throughput, Average Life play, Expected contribution, DUSAF agreements, GL30 agreements, Disagreements.

Game Evaluation

Quantity	Value	Unit of measure
N° Of Players	68	#
Total Played Time	0 20:17:08	d HH:mm:ss
N° Of Validated Pixels	1600	#
N° Of Played Pixels	1600	#
Total N° Of Pixels	1600	#
Completion rate	100.00	%
Throughput	78.92	solved tasks/hour
ALP (Average Life Play)	17.90	minutes/player
Expected contribution	23.54	solved tasks/player
DUSAF Agreements	86.82	%
GL30 Agreements	11.87	%
Disagreements	1.31	%

Figure 44: Evaluation of Land Cover Validation game.

The throughput is 78.92 solved tasks/hour, which is a good result as we saw in the previous chapters. The Average Life play is 17.90 minutes/Player, which gives enjoys ability of the game and 17.90 is the compromising results. The expected contribution is 23.54 solved tasks/player which means each player helped in validating 23.54 pixels approximately. The 86.82% of classification made by the players are in agreement with the DUSAF category, 11.87% of classification made by the players is in agreement with GL30 category and 1.31% of classification has been in agreement with the classification which is neither DUSAF nor GL30 category.

6.2.2 User Analysis

The Land Cover Validation game participants were analysed initially. From the analysis, it was found there were more female participants when compared with the male participants (Figure 45).

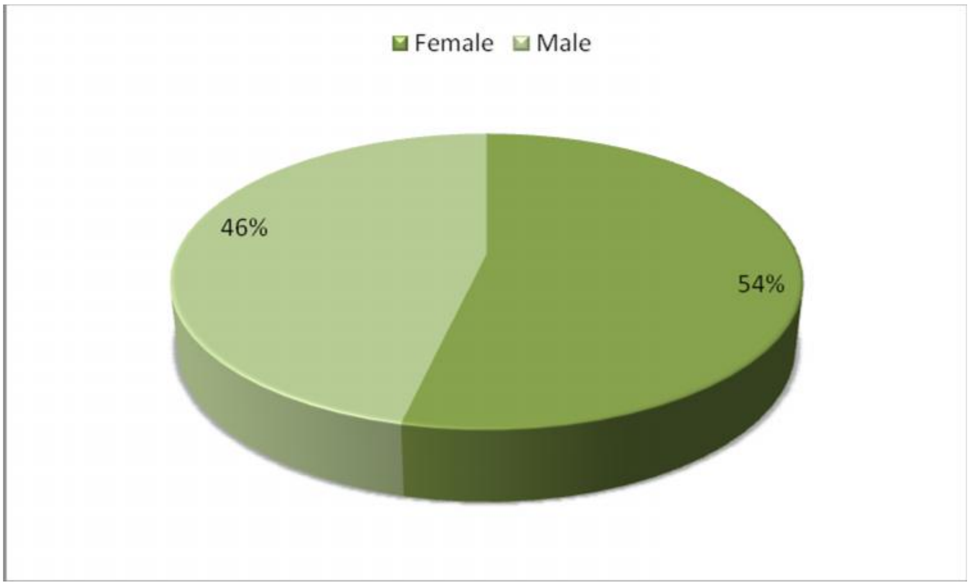


Figure 45: Gender analysis of Land Cover Validation game.

The life play of both gender was computed, surprisingly female life play was found to be more than the male's life play (Figure 46).

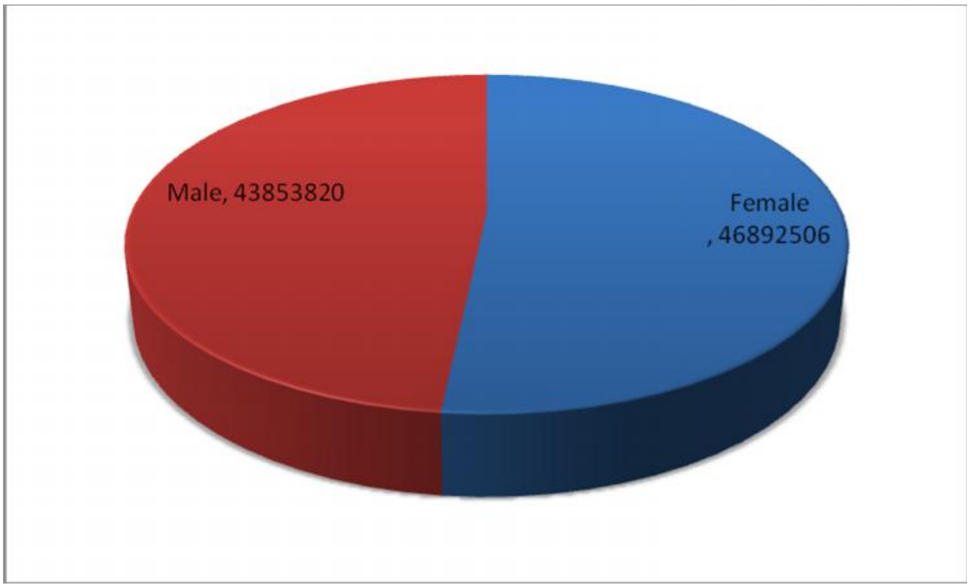
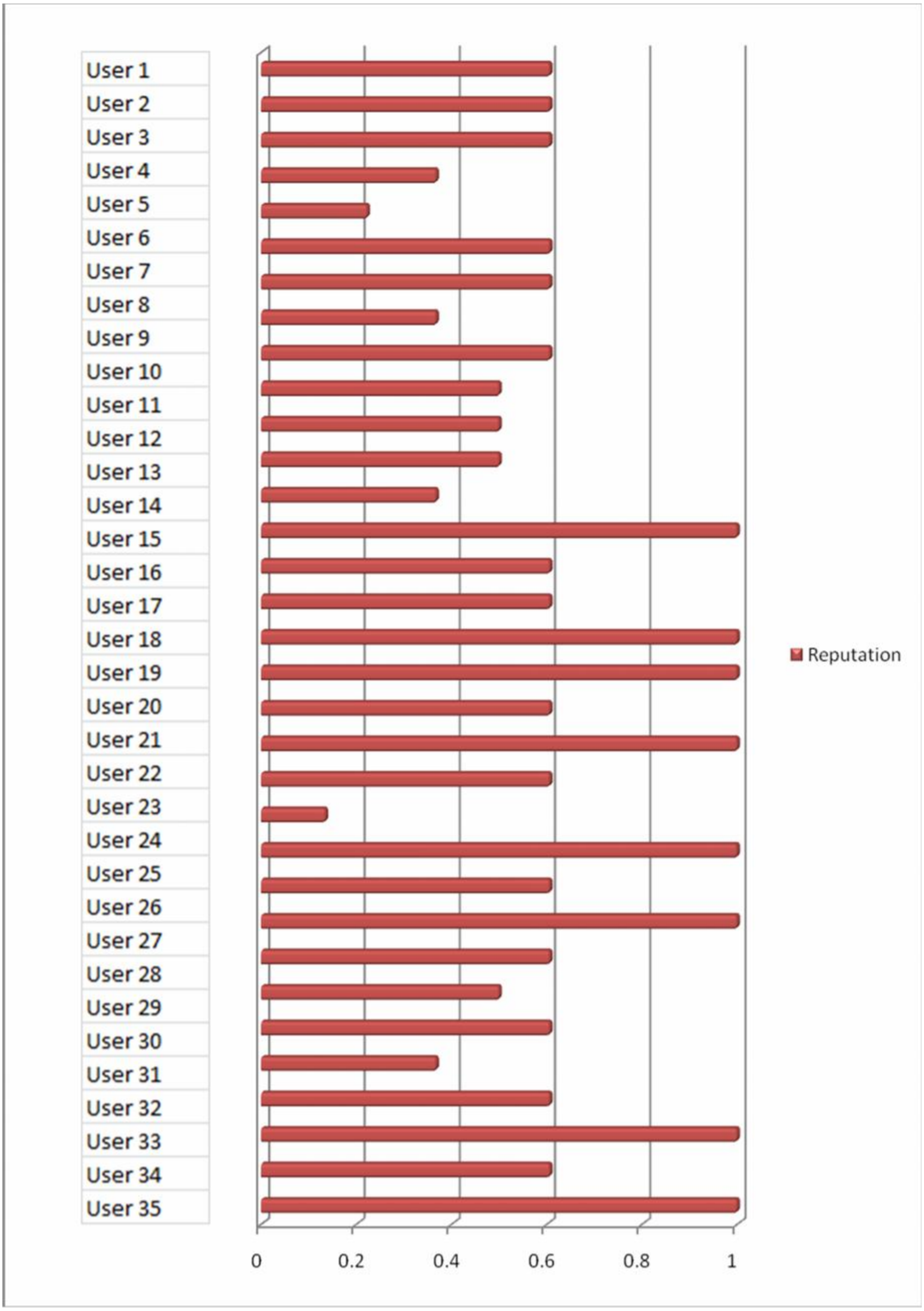


Figure 46: Life play of both genders.

The reputation of each player is computed as we saw in the previous chapter. The figure 47 shows the reputation of each player in the Land Cover Validation game.



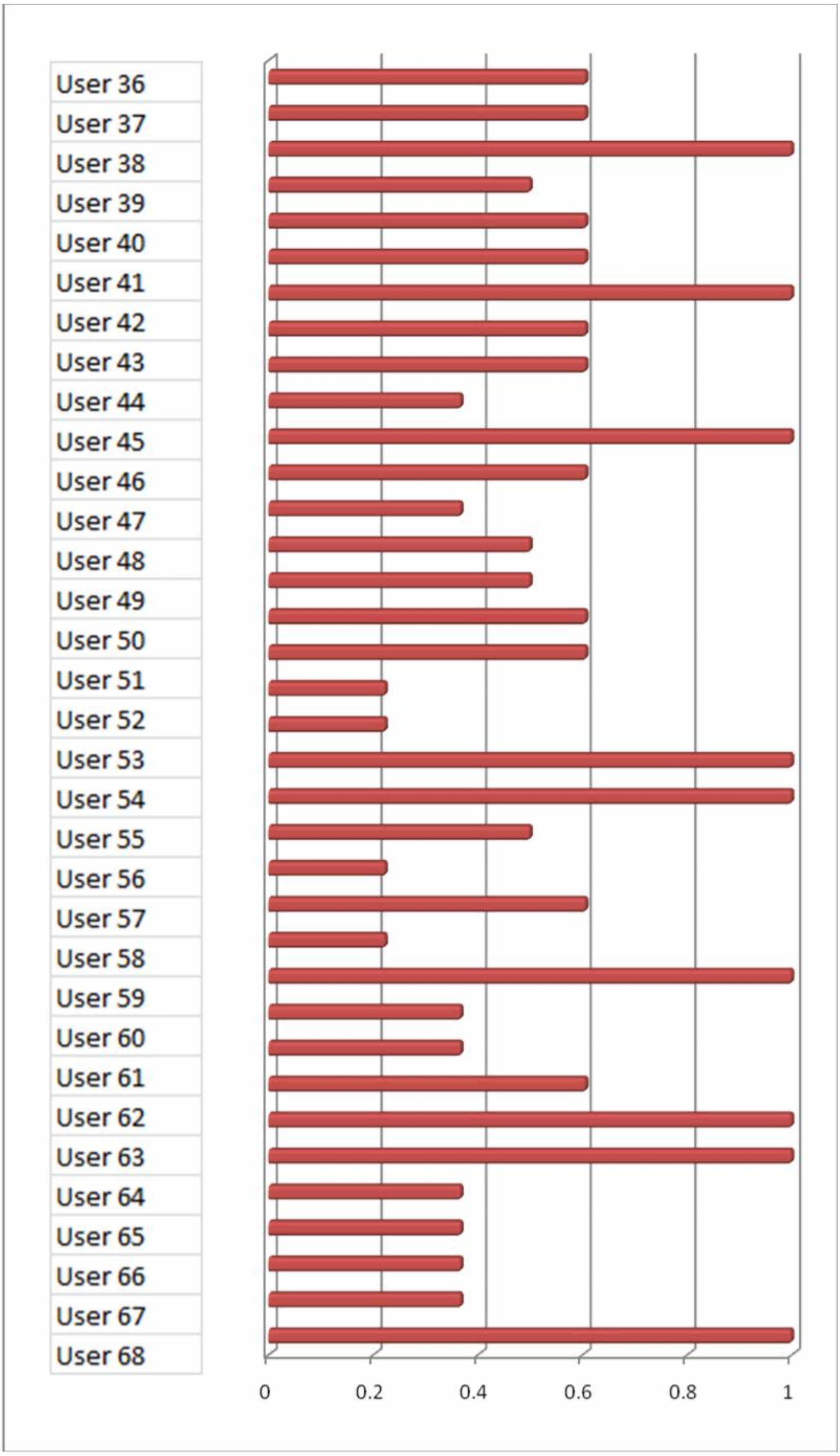


Figure 47: Reputation of each user.

The figure 48 gives the Number of players in the different reputation. From the given figure it is clear that 76% of the players have reputation more than 0.5. 30% players have the maximum reputation. This shows the accuracy of the validation should be high.

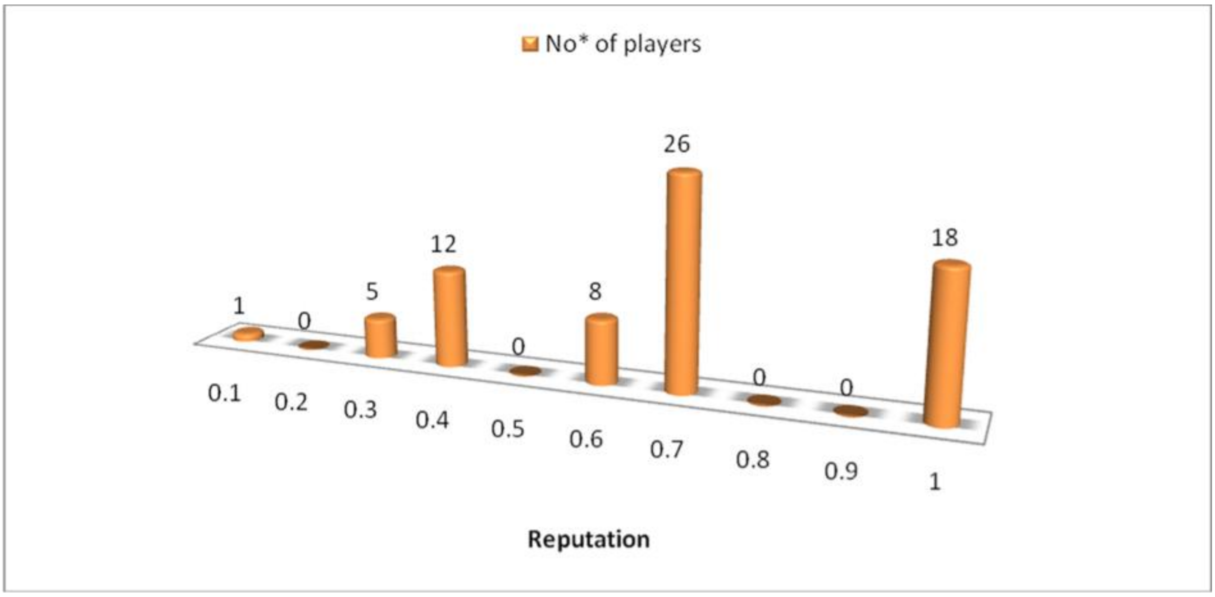


Figure 48: Number of players in different reputation range.

From the user analysis it has been found the maximum user who entered the game used Facebook for the authentication rather than Google+ and Twitter. Twitter was used least (Figure 49).

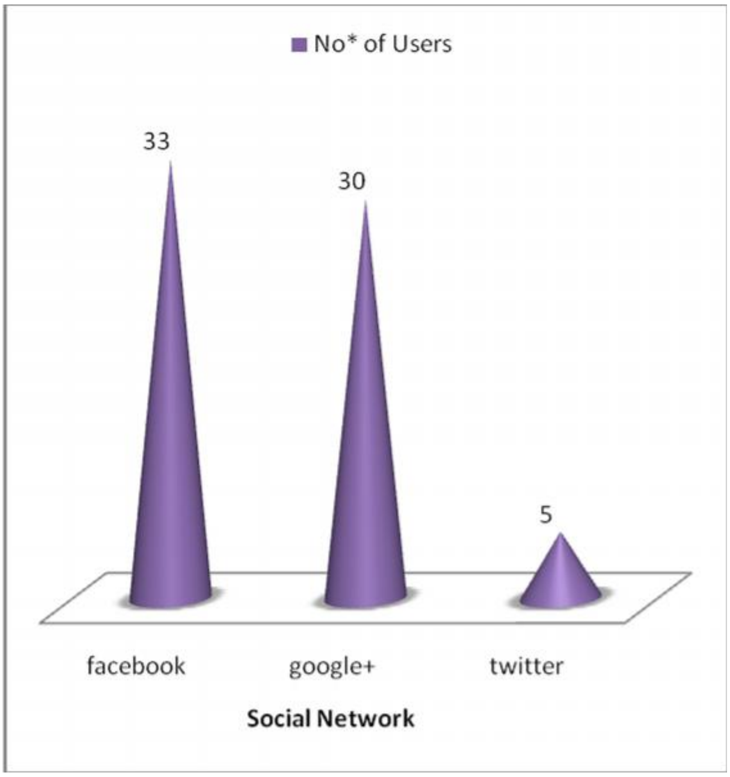


Figure 49: Social Network user analysis graph.

Player	Contribution	Player	Contribution
User1	1190	User34	311
User2	5	User35	10
User3	60	User36	60
User4	5	User37	54
User5	160	User38	130
User6	10	User39	6
User7	12	User40	5
User8	40	User41	172
User9	61	User42	234
User10	20	User43	181
User11	65	User44	18
User12	292	User45	9
User13	1704	User47	1
User14	905	User48	82
User15	1876	User49	10
User16	60	User50	210
User17	3	User51	5
User18	15	User52	100
User19	71	User53	147
User20	39	User54	4
User21	1	User55	4
User22	1	User56	5
User23	20	User57	40
User24	17	User58	435
User25	66	User59	133
User26	145	User60	125
User27	51	User61	19
User28	30	User62	528
User29	45	User63	5
User30	200	User64	10
User31	89	User65	1
User32	3	User66	5
User33	15	User67	10
		User68	30

Table 12: List of players with their contributions[34].

[34] *Total number of classifications*

The table 12 gives the list of players and their contributions (i.e.) number of classification they performed during the game play. About 60% of the players classified more than 50 pixels. Two players classified more than 1500 pixels which mean they both played a full game in those three days. About 15% of the players just played one round and didn't continue further.

6.2.3 Pixel analysis

The date wise analysis was done to understand how many classifications have been performed during the mapping party in FOSS4G and how many classifications was done after the mapping party. The figure 50 gives the details of number of classifications done on different dates. From the graph it is clear that 72% of the classification has been performed during the mapping party (15-17 July).

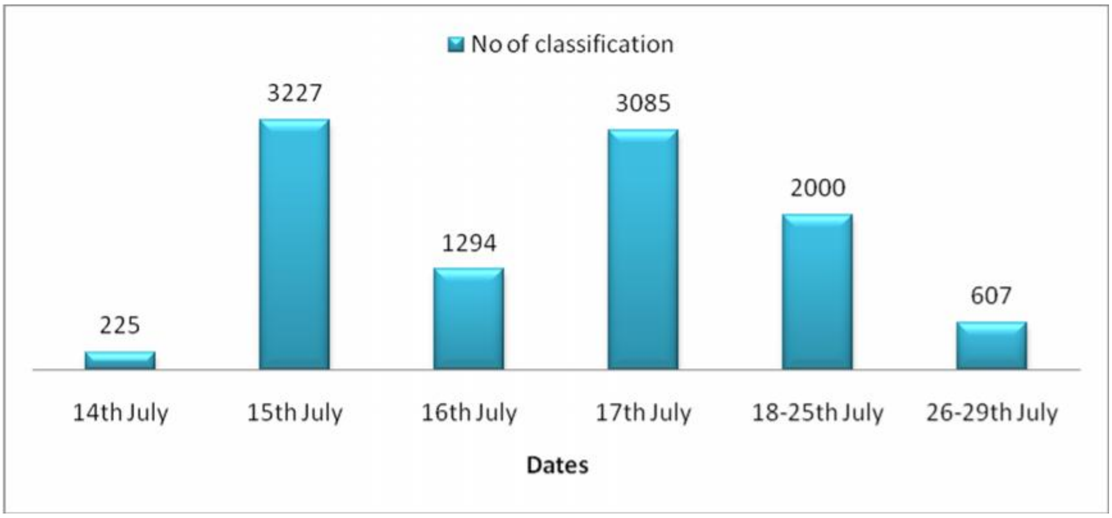


Figure 50: Number of classification for different dates.

Figure 51 shows the number of times the pixels have been classified multiple times. It is clear that 2027 pixels have been classified one to three times, 1891 pixels have been classified four to five times, 1249 pixels have been classified six to nine times, 1106 pixels have been classified ten to fifteen times, 768 pixels have been classified sixteen to twenty times, 657 pixels have been classified twenty-one to thirty times, 395 pixels have been classified thirty-one to forty times, 4 pixels have been classified more than 41 times.

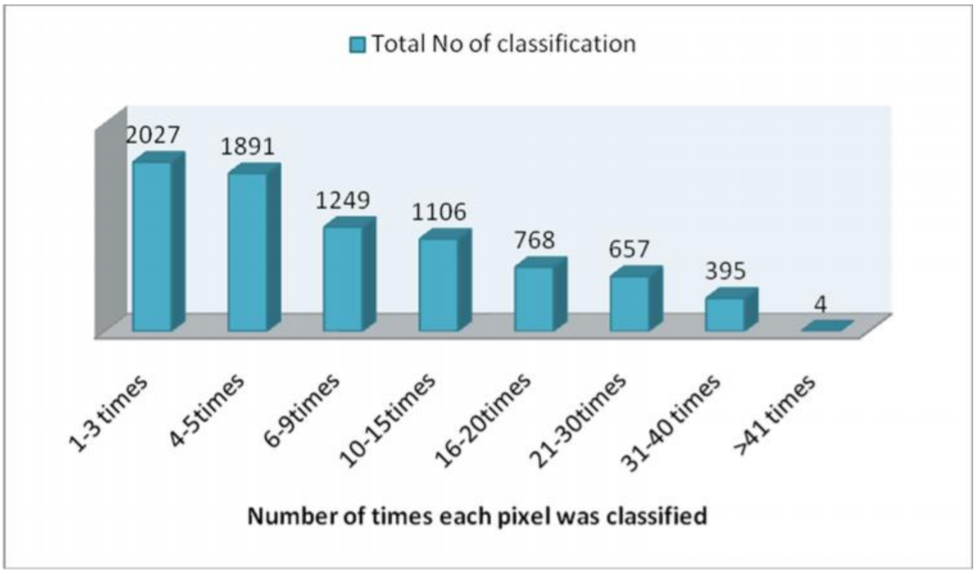


Figure 51: The number of times the pixels have been classified multiple times.

The figure 52 shows the number of times the validated pixels have been classified multiple times.990 validated pixels have been classified one to three times, 310 validated pixels have been classified four to five times, 143 validated pixels have been classified six to nine times, 76validated pixels have been classified ten to fifteen times, 44pixels have been classified sixteen to twenty times, 27validated pixels have been classified twenty-one to thirty times, 7validated pixels have been classified thirty-one to forty times, 4validated pixels have been classified more than 41 times.

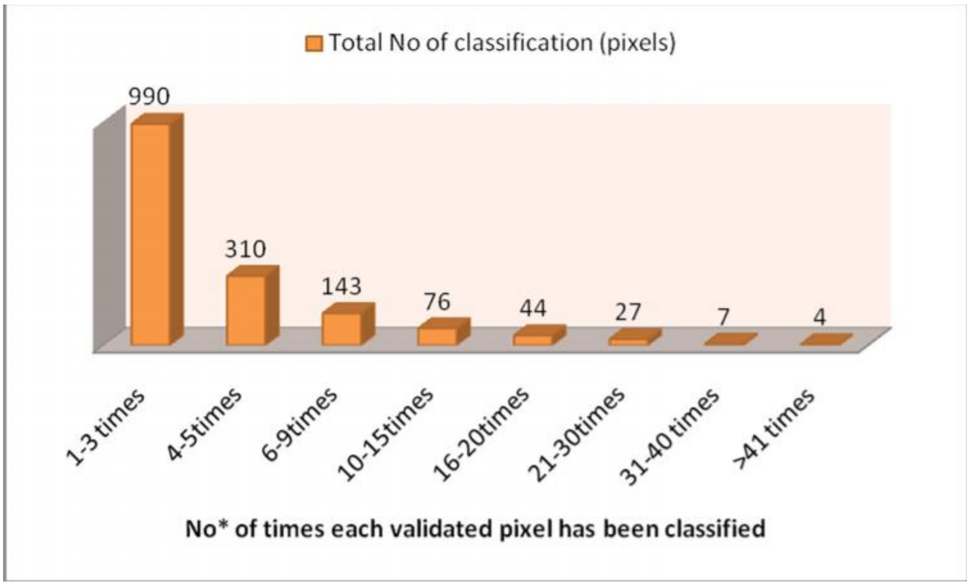


Figure 52: The number of times the validated pixels have been classified many times.

6.2.3.1 Complicated pixels

Figure 53: Contradicting pixel which has been classified by 57 players and validated as DUSAF classification.

The pixel in figure 53 has been classified by 57 players before one category reaches the upper confidence level. For this particular pixel 19 players selected the category artificial areas which correspond to GL30 classification, 27 players selected the category Agricultural area which corresponds to DUSAF classification and 11 players selected the Forest and Semi-Natural area which is neither DUSAF nor GL30 classification. After 57 player's classification this pixel has been validated as an Agricultural area which corresponds to DUSAF.

Figure 54: Contradicting pixel which has been classified by 42 players and validated as DUSAF classification.

The pixel in the figure 54 has been classified by 42 players before one category reaches the upper confidence level. For this particular pixel 19 players selected the category artificial area which corresponds to GL30 classification, 7 players selected the category Agricultural area which is neither GL30 nor DUSAF classification and16 players selected the Forest and Semi-Natural area which corresponds to DUSAF. After 42player's classifications, this pixel has been validated as Forest or Semi-Natural area which corresponds to DUSAF classification.

Figure 55: Contradicting pixel which has been classified by 41 players and validated as DUSAF classification.

The pixel in the figure 55 has been classified by 41 players before one category reaches the upper confidence level. For this particular pixel 18 players selected the category artificial area which corresponds to GL30 classification, 2 players selected the category Agricultural area which is neither GL30 nor DUSAF classification, 18 players selected the Forest and Semi-Natural area which corresponds to DUSAF classification, 1 player selected the Wetland which is neither GL30 nor DUSAF classification and 2 players selected the Water bodies category which is neither GL30 nor DUSAF classification. After 41player's classifications, this pixel has been validated as Forest and Semi-Natural area which corresponds to DUSAF classification.

Figure 56: Contradicting pixel which has been classified by 40 players and validated as DUSAF classification.

The pixel in the figure 56 has been classified by 40 players before one category reaches the upper confidence level. For this particular pixel 13 player selected the category artificial area which corresponds to GL30 classification, 9 players selected the category Agricultural area which is neither GL30 nor DUSAF classification and 18 players selected the Forest and Semi-Natural area which corresponds to DUSAF. After 40 player's classifications, this pixel has been validated as Forest or Semi-Natural area which corresponds to DUSAF classification.

Figure 57: Contradicting pixel which has been classified by 40 players and validated as DUSAF classification.

The pixel in the figure 57 has been classified by 40 players before one category reaches the upper confidence level. For this particular pixel 20 player selected the category artificial area which corresponds to GL30 classification, 1 players selected the category Agricultural area, which is neither GL30 nor DUSAF classification and 19 players selected the Forest and Semi-Natural area which corresponds to DUSAF. After 40 player's classifications, this pixel has been validated as Forest or Semi-Natural area which corresponds to DUSAF classification.

Figure 58: Contradicting pixel which has been classified by 37 players and validated as GL30 classification.

The pixel in the figure 58 has been classified by 37 players before one category reaches the upper confidence level. For this particular pixel 12 player selected the category artificial area which corresponds to DUSAF classification, 4 players selected the category Agricultural area which is neither GL30 nor DUSAF classification and 21 players selected the Forest and Semi-Natural area which corresponds to GL30 classification. After 37 player's classifications, this pixel has been validated as Forest or Semi-Natural area which corresponds to GL30 classification.

Figure 59: Contradicting pixel which has been classified by 33 players and validated as GL30 classification.

The pixel in the figure 59 has been classified by 33 players before one category reaches the upper confidence level. For this particular pixel 7 player selected the category artificial area which corresponds to DUSAF classification, 10 players selected the category Agricultural area which is neither GL30 nor DUSAF classification and 16 players selected the Forest and Semi-Natural area which corresponds to GL30 classification. After 33 player's classifications, this pixel has been validated as Forest or Semi-Natural area which corresponds to GL30 classification.

6.2.3.2 Disagreement pixels

Figure 60: Contradicting pixel which has been classified by 39 players and validated as classification which is neither DUSAF nor GL30.

The pixel in the figure 60 has been classified by 39 players before one category reaches the upper confidence level. For this particular pixel 12 players selected the category artificial area which corresponds to GL30 classification, 10 players selected the category Agricultural area which corresponds to DUSAF classification, 15 players selected the Forest and Semi-Natural area which is neither GL30 nor DUSAF classification and 2 players selected Wetland which is neither GL30 nor DUSAF classification. After 39 player's classifications, this pixel has been validated as Forest or Semi-Natural which is neither GL30 nor DUSAF classification.

Figure 61: Contradicting pixel which has been classified by 44 players and validated as category which is neither DUSAF nor GL30.

The pixel in the figure 61 has been classified by 44 players before one category reaches the upper confidence level. For this particular pixel 2 players selected the category artificial area which is neither GL30 nor DUSAF classification, 10 players selected the category Agricultural area which corresponds to GL30 classification, 21 players selected the Forest and Semi-Natural area which is neither GL30 nor DUSAF classification, 10 players selected the Wetland which corresponds to DUSAF classification and 1 player selected the Water bodies category which is neither GL30 nor DUSAF classification. After 44 players classifications this pixel has been validated as Forest and Semi-Natural area which is neither GL30 nor DUSAF classification.

The figure 62 & 63 displays all the validated pixels that doesn't belong to neither DUSAF nor GL30 classification. There were 21 such disagreements pixels which has been provided in the initial result of Game evaluation (Figure 44)

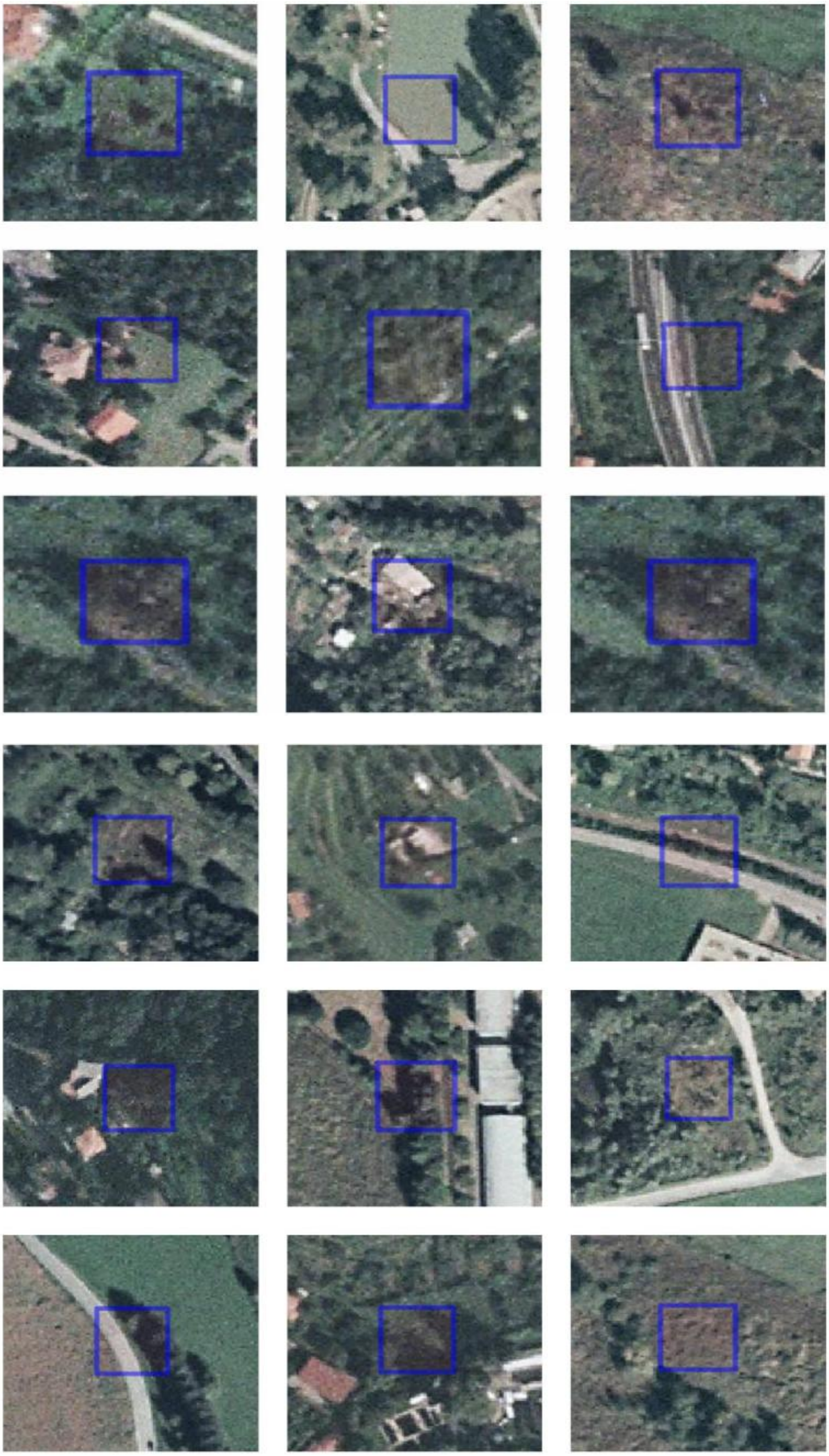


Figure 62: The pixels that were validated as category which is neither DUSAF nor GL30.

Figure 63: The pixels that were validated as category which is neither DUSAF nor GL30.

6.2.3.3 Non-Validated pixels

Some of the pixel still exists in the game which is not yet validated because of the contradictory classification between the players. Some of such pixels have been extracted in order to analyse further. The figure 64 shows the list of pixel which are not yet validated.

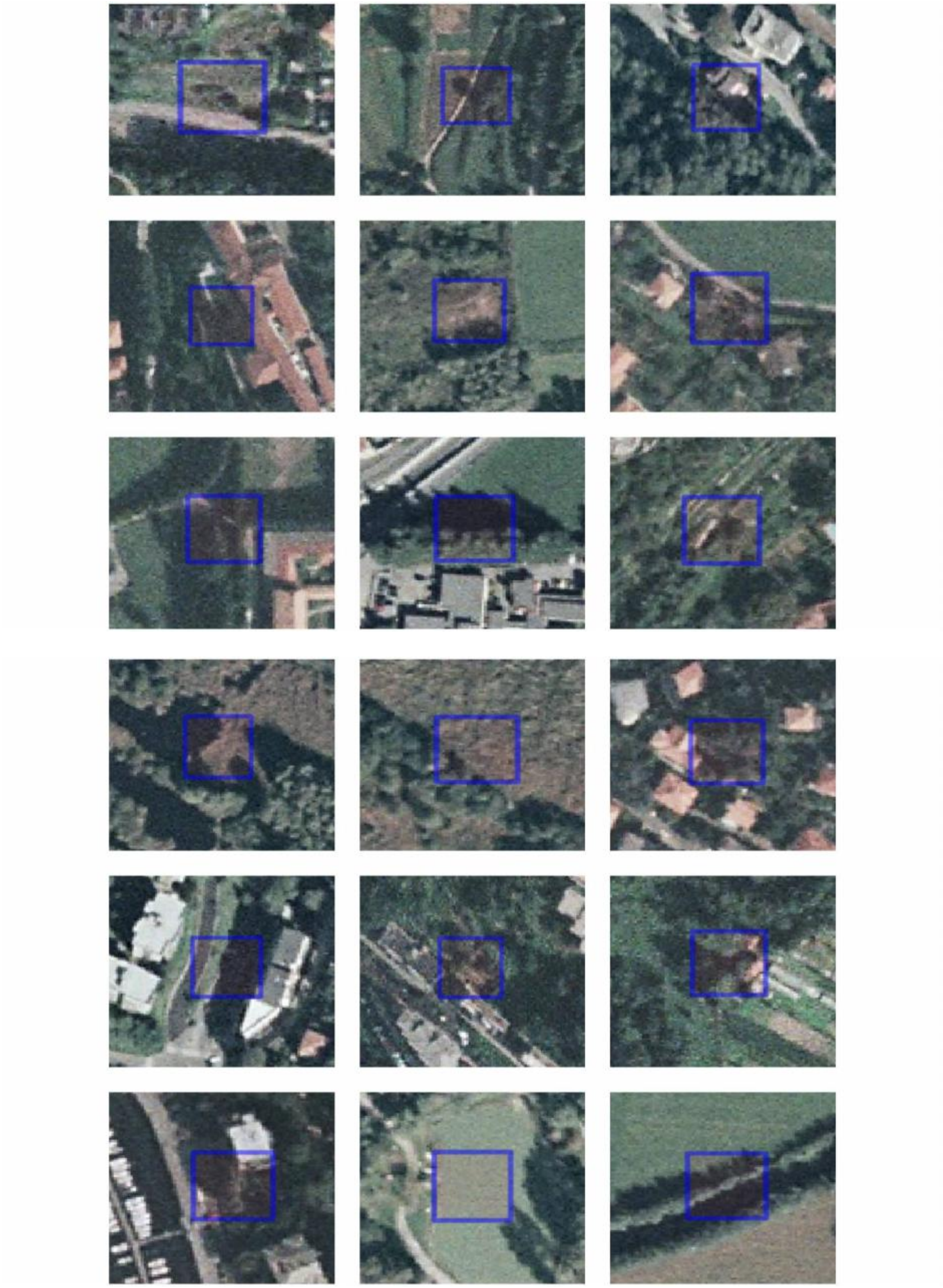


Figure 64: The example of pixel which are not yet validated and still exists in the game.

Chapter 7

7. Conclusion and Future Improvements

The most important component of the land cover classification process is accuracy assessment. The maps without the accuracy associated with them remain as untested hypotheses. The statistically valid estimates of map accuracy and their publication are essential for validation of land cover products and their ultimate acceptance and use. A set of core analysis methods exists for accuracy assessment that should be routinely adopted as a baseline for reporting map accuracy. These methods include employing probability sampling and consistent estimators within the design-based inference framework to generate estimates of the overall accuracy of the map as well as per-class accuracies and the variances of this estimate. Confusion matrices should be presented with the accuracy assessment, and the data used to derive these estimates should be archived and made accessible to the scientific community.

In this research work, we assessed the accuracy by employing sampling and consistent estimators within the design-based inference framework which includes overall accuracy, producer accuracy, user accuracy, allocation and quantity agreement. This research work didn’t stop with the basic level of accuracy assessment. It also extended to another critical feature of global validation, which is the reference data that is collected from the common underlying sampling design. The sampling design must have the capacity to increase the sample size for rare land-cover classes. In this research work, cell by cell comparison has been preformed. Another level of validation has been performed for the disagreement pixel based on the human computational approach using crowd sourcing.

The Land Cover Validation game presented here allows end-users with minimal training in image processing to classify the land-cover by accessing orthophoto. The finer-resolution assessments of land-cover as attempted using this game, allows large-scale land cover maps to be validated. The adopted architecture minimises the need to exchange large sets of data over the internet by focusing in a defined area and allows the players to be independent from the working environment, since only a web browser is required. The current implemented game does not promise to replace the current land cover validation activities by experts, but for sure provides the additional data to improve the quality assurance of the validation process.

The land cover validation game gives answers to many different research questions such as

1. How hybrid land cover products are obtained from different types of data integration?
2. What kind of novelty methods are implemented in creating the hybrid products?
3. How do we create a sustainable community engagement campaign to validate the data products?

4. How to attract ordinary citizens to engage in the land cover validation process?
5. Which classification method has maximum percentage of agreement by the crowd?
6. What percent of the classification agreed with the classification which is neither GL30 nor DUSAF?
7. Which innovative method is adopted to classify the contributors (players) of validation process?

Looking at the results of the game, the land cover lass semantics are potentially the largest influence on DUSAF agreement, rather than GL30 or other classification. These categorical land cover classification tasks are difficult. Since the first level of CORINE nomenclature has been used, it would be easier for the players to classify them without complexity. A possible explanation for why the first level nomenclature has been used, is that this categorical classification scheme is generalizing land covers too much, and these classes are at too high of a level that subjectivity overrides objectivity. The earth's surface is complex and heterogeneous; hence increasing the complexity into relatively high-level categorical classes is prone to errors and disagreements.

7.1 Lesson learned and future improvement

Though the reputation of the users is not bad, it is necessary to develop a new online community which comprises of remote sensing experts with an interest in land cover or the broader public who need to engage on a broader environmental level. At present, the system is designed in such a way that it promotes one-way communication, i.e. the players provide assessments of the land cover, but they cannot view the results or they cannot view their status of the validation. Also the player does not receive any feedback for instance, the effect of providing the information in terms of potential improvements to land cover, or any social interaction with other users or the scientist on the Land Cover Validation team.

As part of the developing community, each player's classification and their agreements should be made available in the game (Figure 65), initiating such things guide the crowd in a type of collaborative learning process. Through a collaborative land cover mapping efforts, groups could work together or individuals that provide land cover assessments of the same area could discuss their findings, particularly in the areas, which are difficult to classify because of low visibility, heterogeneous landscape and low resolution imagery. The individuals stimulate their learning experience by involving with the scientist in such collaborative work. A feature will be added that allows the players to display their past contributions, such that those pixels can be played again and be corrected. These types of elements are crucial if ordinary citizens are to be engaged.

Player Contribution

Total N° Of Pixels	1600	#
Completion rate	0.13	%
N° Of Played Pixels	147	#
N° Of Validated Pixels	2	#
E: Total Played Time	0 00:19:11	d HH:mm:ss
DUSAF Agreements	100.00	%
GL30 Agreements	0.00	%
Disagreements	0.00	%

Figure 65: Player Contribution page.

The new method can be built to different the players background such that the evaluation can be done according to their background. The implementation of the game in Facebook games might attract many users with different background but results obtained from them may not useful for the scientific evaluation. But implementing the game specific to Geospatial community or land cover experts group fetch better results than the one which is mentioned before. So it would be nice to have a method to differentiate various users. The experimentation with data quality is currently ongoing by making use of experts. Some of the implementation which fetches better results is as follows.

1. Systematic analysis of contributions by experts.
2. Implementation of better internal rating system than reputation is needed to indicate the degree to which the contributions can be trusted.
3. Introducing the players with common forum based on specific area to discuss and assess the land cover can be a trusted way to improve the quality.
4. Getting the feedback from the players would be very useful.
5. If the users could define their own legend classes in their own language, this kind of user-defined legend can be useful in assessing the current global land cover product based on these legends or more useful to assess maps supplied by the user through a WMS or WCS.

The process of analysing the data for the player performance has been already implemented in the game but still the confidence level to eliminate the pixel from the game has to be improved in order to obtain the definite value for the confidence level. The validated product can be integrated with many existing land cover to produce a single, improved product. The resulting map from the validation game can be used for different applications. Incorporating the Flickr photos into the game makes the validation more accurate, as you can see in the figure 66. The player while playing can verify their answers with the Flickr image which is Geo-coded into the game. In some cases where the Flickr images or similar ones are not available, the player has to select his own choice.

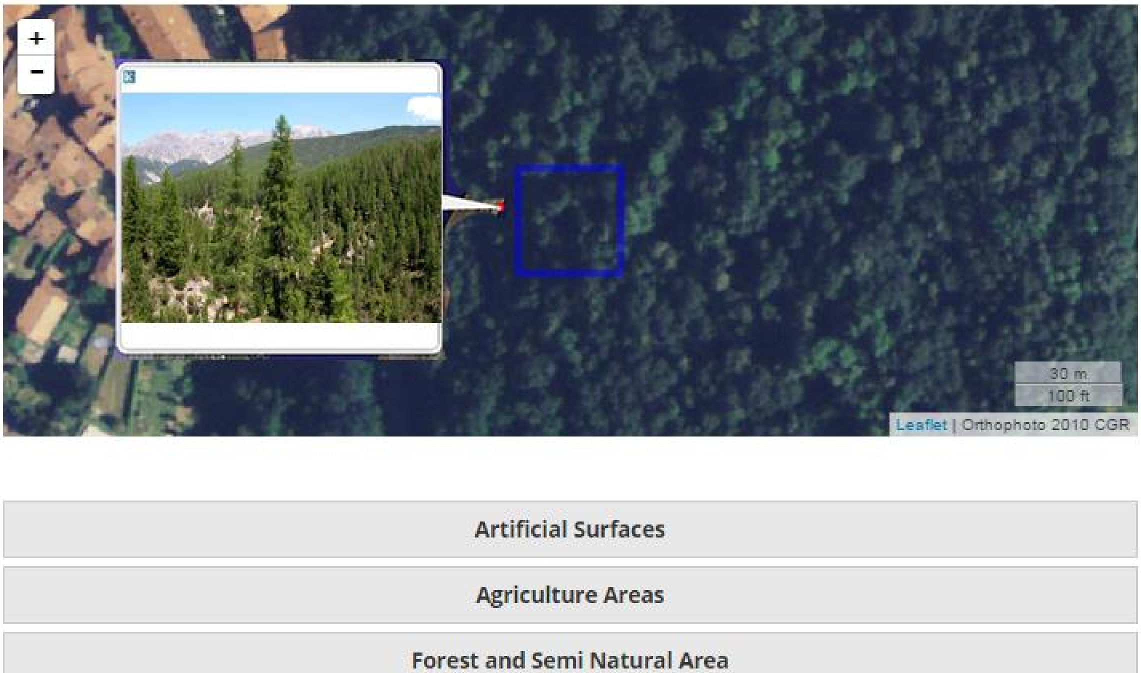


Figure 66: Incorporating the Flickr photos into the Land Cover Validation game.

Bibliography

[i]Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F.,& Obersteiner, M. (2009). Geo-Wiki. Org: The use of crowdsourcing to improve global land cover. Remote Sensing, 1(3), 345-354.

[ii] Bastin, L., Buchanan, G., Beresford, A., Pekel, J.F., & Dubois, G. (2013). Open-source mapping and services for Web-based land-cover validation. Ecological Informatics, 14, 9-16.

[iii] Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., & Mills, J. (2014). Global land cover mapping at 30 m resolution: A POK-based operational approach. ISPRS J. Photogram. Remote Sensing. http://dx.doi.org/10.1016/j.isprsjprs.2014.09.002.

[iv] Brovelli, M.A., Molinari, M.E., Hussein, E., Chen, J., & Li, R. (2015). The first comprehensive accuracy assessment of GlobeLand30 at a national level: methodology and results. Remote Sensing, 7, 4191-4212. doi:10.3390/rs70404191.

[v] Olofsson, P., Stehman,S.V., WoodcockC.E., Sulla-Menashe,D., Sibley,A M., Newell, J,D., Mark, A, Friedl., & Martin,H. A global land-cover validation data set, part I: fundamental design principles. International Journal of Remote Sensing. Volume 33, Issue 18, 2012.

[vi]McIver, D. K. and M. A. Friedl (2001). "Estimating pixel-scale land cover classification confidence using non-parametric machine learning methods." IEEE Transactions on Geoscience and Remote Sensing 39(9): 1959-1968.

[vii] Sparks, K., Klippel, A., Wallgrün, J. O., Mark, D. M., Bertolotto, C. D., Freundschuh, S. M., & Bell, S. (2011). Citizen Science Land Cover Classification Based on Ground and Aerial Imagery.

[viii] Nusser, S. M., & Goebel, J. J. (1997). The National Resources Inventory: a long-term multi-resource monitoring programme. Environmental and Ecological Statistics, 4(3), 181-204.

[ix] Strahler, A. H., Boschetti, L., Foody, G. M., Friedl, M. A., Hansen, M. C., Herold, M., ... & Woodcock, C. E. (2006). Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps. European Communities, Luxembourg, 51.

[x] Perger, C., Fritz, S., See, L., Schill, C., VAN DER VELDE, M., MCCALLUM, I., & Obersteiner, M. (2012). A campaign to collect volunteered geographic Information on land cover and human impact. GI_Forum, 83-91.

[xi]See, L., Comber, A., Salk, C., Fritz, S., van der Velde, M., Perger, C., ... & Obersteiner, M. (2013). Comparing the quality of Crowdsourced data contributed by expert and non-experts. PloS one, 8(7).

[xii] Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C., & Boyd, D. S. (2013). Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project. Transactions in GIS, 17(6), 847-860.

[xiii]Comber, A., Brunsdon, C., See, L., Fritz, S., & McCallum, I. (2013). Comparing Expert and Non-expert Conceptualisations of the Land: An Analysis of Crowdsourced Land Cover Data. In Spatial Information Theory (pp. 243-260). Springer International Publishing.

[xiv]Comber, A., See, L., & Fritz, S. (2014). The Impact of Contributor Confidence, Expertise and Distance on the Crowdsourced Land Cover Data Quality. GI_Forum 2014-Geospatial Innovation for Society.

[xv]Iwao, K., Nishida, K., Kinoshita, T., & Yamagata, Y. (2006). Validating land cover maps with Degree Confluence Project information. Geophysical Research Letters, 33(23).

[xvi] Foody, G. M., & Boyd, D. S. (2013). Using volunteered data in land cover map validation: Mapping West African forests. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, 6(3), 1305-1312.

[xvii]Jokar Arsanjani, J., Helbich, M., Bakillah, M., Hagenauer, J., & Zipf, A. (2013). Toward mapping land-use patterns from volunteered geographic information. International Journal of Geographical Information Science, 27(12), 2264-2278.

[xviii] Prestopnik, N. R., & Crowston, K. (2011, December). Gaming for (citizen) science: exploring motivation and data quality in the context of crowdsourced science through the design and evaluation of a social-computational system. In e-Science Workshops (eScienceW), 2011 IEEE Seventh International Conference on (pp. 28-33). IEEE.

[xix]Irwin, A. (1995). Citizen science: A study of people, expertise and sustainable development. Psychology Press.

[xx]Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioScience, 59(11), 977-984.

[xxi] Roy, H. E., Pocock, M. J. O., Preston, C. D., Roy, D. B., Savage, J., Tweddle, J. C., & Robinson, L. D. (2012). Understanding citizen science and environmental monitoring: final report on behalf of UK Environmental Observation Framework.

[xxii]Rowland, K. (2012). Citizen science goes' extreme'. Nature News, 17.

[xxiii]Snäll, T., Kindvall, O., Nilsson, J., & Pärt, T. (2011). Evaluating citizen-based presence data for bird monitoring. Biological conservation, 144(2), 804-810.

[xxiv]Haklay, M. (2012). Citizen Science and Volunteered Geographic Information – overview and typology of participation in Sui, D.Z., Elwood, S. and M.F. Goodchild (eds.), 2012. Crowdsourcing Geographic Knowledge: Volunteered Geographic Information(VGI) in Theory and Practice. Berlin: Springer. pp 105-122. DOI: 10.1007/978-94-0074587-2_7.

[xxv]Rosner, H. (2013). Data on wings. Scientific American, 308(2), 68-73.

[xxvi] Howe, J. (2010). Crowdsourcing: Why the power of the crowd is driving the future of business. [Online]. Available: www.crowdsourcing.com/ [Accessed: 22nd September 2013].

[xxvii]Moore, K. (2011). Survey: How Much Would You Pay to Save the Bay? Barnegat Bay Partnership. [Online]. Available: http://bbp.ocean.edu/pages/109.asp?item=584 [Accessed: 25th July 2013].

[xxviii]Graham, E. A., Henderson, S., & Schloss, A. (2011). Using mobile phones to engage citizen scientists in research. Eos, Transactions American Geophysical Union, 92(38), 313-315.

[xxix]Wiggins, A., & Crowston, K. (2011, January). From conservation to crowdsourcing: A typology of citizen science. In System Sciences (HICSS), 2011 44th Hawaii international conference on (pp. 1-10). IEEE.

[xxx]Grey, F. (2009). Viewpoint: The age of citizen cyberscience. CERN Courier, 29th April. [Online]. Available: http://cerncourier.com/cws/article/cern/38718 [Accessed: 28th October 2013].

[xxxi] Gollan, J., de Bruyn, L. L., Reid, N., & Wilkie, L. (2012). Can volunteers collect data that are comparable to professional scientists? A study of variables used in monitoring the outcomes of ecosystem rehabilitation. Environmental management, 50(5), 969-978.

[xxxii] Gardiner, M. M., Allee, L. L., Brown, P. M., Losey, J. E., Roy, H. E., & Smyth, R. R. (2012). Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. Frontiers in Ecology and the Environment, 10(9), 471–476. DOI:10.1890/110185

[xxxiii]Rowland, K. (2012). Citizen science goes “extreme”. Nature, 17th February. DOI:10.1038/nature.2012.1005

[xxxiv]Grove-White, R., Waterton, C., Ellis, R., Vogel, J., Stevens, G., & Peacock, B. (2007). Amateurs as Experts: Harnessing New Networks for Biodiversity: End of Award Report. 1–179. Lancaster University and Natural History Museum.
Available: http://csec.lancs.ac.uk/docs/Amateurs%20as%20Experts%20Final%20Report.pdf [Accessed: 24th October 2013].

[xxxv]Shirk, J.L., Ballard, H.L., Wilderman, C.C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., Minarchek, M., Lewenstein, B.V., Krasny, M.E. and Bonney, R. (2012). Public participation in scientific research: a framework for deliberate design. Ecology and Society, 17(2), 29-48. DOI: 10.5751/ES-04705-170229

[xxxvi] Tweddle, J.C., Robinson, L.D., Pocock, M.J.O. & Roy, H.E (2012). Guide to citizen science: developing, implementing andevaluating citizen science to study biodiversity and the environment in the UK. Natural History Museum and NERC Centre for Ecology & Hydrology for UKEOF. 1-36. [Online]. Available: www.ceh.ac.uk/products/publications/documents/citizenscienceguide.pdf [Accessed:24th July 2013]

[xxxvii]Mathieson, K. (2013). The 3Rs: Citizen science in the classroom. British Science Association. [Online]. Available: www.britishscienceassociation.org/blog/3rs-citizen-science-classroom [Accessed: 24th July 2013].

[xxxviii]Delaney, D. Sperling, C.D., Adams,C.S., Leung, B. (2008). Marine invasive species: validation of citizen science and implications for national monitoring networks. Biological Invasions, 10(1), 117-128. DOI: 10.1007/s10530-007-9114-0

[xxxix]Von Ahn, L. (2006). Games with a purpose. Computer, 39(6), 92-94.

[xl] Prestopnik, N. R., & Crowston, K. (2011, December). Gaming for (citizen) science: exploring motivation and data quality in the context of crowdsourced science through the design and evaluation of a social-computational system. In e-Science Workshops (eScienceW), 2011 IEEE Seventh International Conference on (pp. 28-33). IEEE.

[xli] Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: defining gamification. In Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (pp.9-15). ACM.

[xlii]Von Ahn, L., & Dabbish, L. (2008). Designing games with a purpose. Communications of the ACM, 51(8), 58-67.

[xliii]Sawyer, B., & Smith, P. (2008, February). Serious games taxonomy. In Slides from the Serious Games Summit at the Game Developers Conference.

[xliv]Bjork, S., & Holopainen, J. (2006). Games and design patterns. The game design reader, 410-437.

[xlv]McGonigal, J. (2011). Reality is broken: Why games make us better and how they can change the world. Penguin

[xlvi] Hunicke, R., LeBlanc, M., & Zubek, R. (2004, July). MDA: A formal approach to game design and game research. In Proceedings of the AAAI Workshop on Challenges in Game AI (Vol. 4).

[xlvii]Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction*. Cognitive science, 5(4), 333-369.

[xlviii]Belman, J., & Flanagan, M. (2010). Designing games to foster empathy. International Journal of Cognitive Technology, 15(1), 11.

[xlix]Townshend, J., Justice, C., Li, W., Gurney, C., & McManus, J. (1991). Global land cover classification by remote sensing: present capabilities and future possibilities. Remote Sensing of Environment, 35(2), 243-255.

[l] Credali, M., Fasolini, D., Minnella, L., Pedrazzini, L., Peggion, M., & Pezzoli, S. (2011). Tools for territorial knowledge and government. In D. Fasolini, S. Pezzoli, V. M. Sale, M. Cesca, S. Coffani, & S. Brenna (Eds.), Land cover changes in Lombardy over the last 50 years (pp.17-19). Milano, ERSAF.

[li]Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., & Mills, J. (2014). Global land cover mapping at 30 m resolution: A POK-based operational approach. ISPRS J. Photogram. Remote Sensing. http://dx.doi.org/10.1016/j.isprsjprs.2014.09.002.

[lii]Foody, G. M. (2011). Impacts of imperfect reference data on the apparent accuracy of species presence–absence models and their predictions. Global Ecology and Biogeography, 20(3), 498-508.

[liii]Anderson, J.R.; Hardy, E.E; Roach, J.T.; Witmer, R.E. A land Use and Land Cover Classification System for Use with Remote Sensor Data; Geological Survey Professional Paper 964; U.S. Geological Survey: Reston, VA, USA, 1976.

[liv]Pringle, M.J.; Schmidt, M.; Muir, J.S. Geostatistical interpolation of SLC-off Landsat ETM+ images. ISPRS J. Photogramm. Remote Sens. 2009, 64, 654–664.

[lv]Thomlinson, J.R.; Bolstad, P.V.; Cohen, W.B. Coordinating methodologies for scaling land cover classifications from site-specific to global: Steps toward validating global map products. Remote Sens. Environ. 1999, 70, 16–28.

[lvi]Rosenfield, G. H., & Fitzpatrick-Lins, K. (1986). A coefficient of agreement as a measure of thematic classification accuracy. Photogrammetric engineering and remote sensing, 52(2), 223-227.

[lvii]Congalton, R. G., Oderwald, R. G., & Mead, R. A. (1983). Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. Photogrammetric Engineering and Remote Sensing.

[lviii]Foody, G. M., Campbell, N. A., Trodd, N. M., & Wood, T. F. (1992). Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification. Photogrammetric Engineering and Remote Sensing, 58(9), 1335-1341.

[lix]Stehman, S. V., & Czaplewski, R. L. (1998). Design and analysis for thematic map accuracy assessment: fundamental principles. Remote Sensing of Environment, 64(3), 331-344.

[lx]Pontius Jr, R. G., & Millones, M. (2008). Problems and solutions for kappa-based indices of agreement. Studying, Modeling and Sense Making of Planet Earth, Mytilene, Greece.

[lxi]Stehman, S. V. (2000). Practical implications of design-based sampling inference for thematic map accuracy assessment. Remote Sensing of Environment, 72(1), 35-45.

[lxii]Cochran, W. G. (1977). Sampling techniques (3rd ed.). New York: John Wiley & Sons.

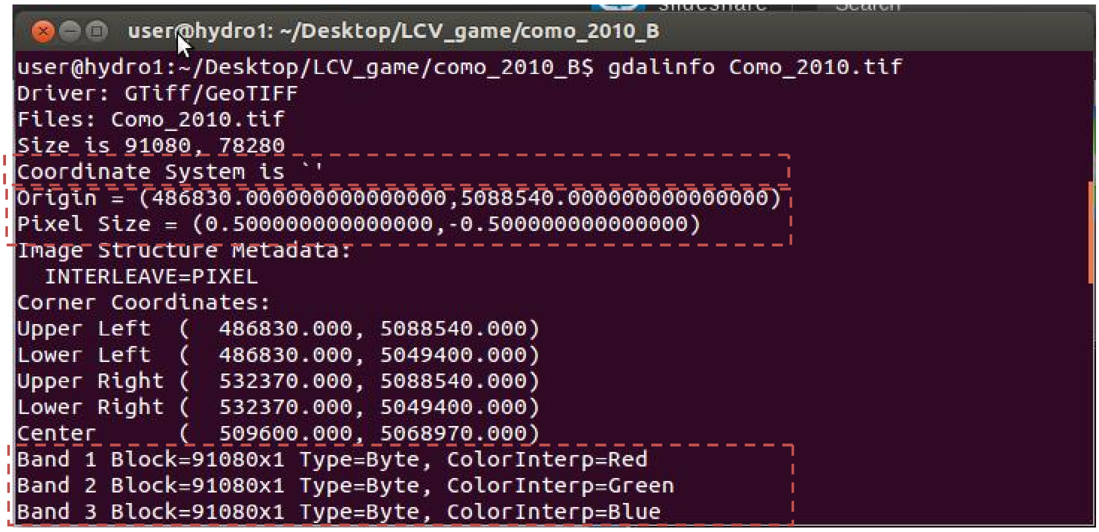
[lxiii] Celino, I., Contessa, S., Corubolo, M., Dell'Aglio, D., Della Valle, E., Fumeo, S., & Krüger, T. (2012). Linking smart cities datasets with human computation–the case of urbanmatch. In The Semantic Web–ISWC 2012 (pp. 34-49). Springer Berlin Heidelberg.

[lxiv]Law, E., & Ahn, L. V. (2011). Human computation. Synthesis Lectures on Artificial Intelligence and Machine Learning, 5(3), 1-121.

Appendix

A. Preprocessing of Orthophotos and tuning Geoserver

The problems (listed in section 5.3.5) which have been faced during the testing phase were rectified by analysing the GDAL information. The figure A-1 highlights the missing CRS information, bad tiling information, and georeferencing information.



```
user@hydro1: ~/Desktop/LCV_game/como_2010_B
user@hydro1:~/Desktop/LCV_game/como_2010_B$ gdalinfo Como_2010.tif
Driver: GTiff/GeoTIFF
Files: Como_2010.tif
Size is 91080, 78280
Coordinate System is `'
Origin = (486830.000000000000000,5088540.000000000000000)
Pixel Size = (0.500000000000000,-0.500000000000000)
Image Structure Metadata:
  INTERLEAVE=PIXEL
Corner Coordinates:
Upper Left  (  486830.000, 5088540.000)
Lower Left  (  486830.000, 5049400.000)
Upper Right (  532370.000, 5088540.000)
Lower Right (  532370.000, 5049400.000)
Center      (  509600.000, 5068970.000)
Band 1 Block=91080x1 Type=Byte, ColorInterp=Red
Band 2 Block=91080x1 Type=Byte, ColorInterp=Green
Band 3 Block=91080x1 Type=Byte, ColorInterp=Blue
```

Figure 67: GDAL info showing missing CRS, Georeferencing and bad Tiling.

STEP 1: The inner tiling of the raster data, setting CRS and Geo-Referencing can be performed using the functional codes as follows

gdal_translate -a_srs "EPSG:32632" -co "TILED=YES" -co "BLOCKXSIZE=512" -co "BLOCKYSIZE=512" input.tif output.tif

STEP 2: The rebuilding overview image is performed using following function codes

gdaladdo -r average output.tif 2 4 8 16 32

After performing *gdal_translate* and *gdaladdo,* the processed data are set with proper CRS and retailed with 512*512 sized pixels (see figure A-2)

```
user@hydro1: ~/Desktop/LCV_game/como_2010_B
user@hydro1:~/Desktop/LCV_game/como_2010_B$ gdalinfo Como_2010_new.tif
Driver: GTiff/GeoTIFF
Files: Como_2010_new.tif
Size is 91080, 78280
Coordinate System is:
PROJCS["WGS 84 / UTM zone 32N",
    GEOGCS["WGS 84",
        DATUM["WGS_1984",
            SPHEROID["WGS 84",6378137,298.257223563,
                AUTHORITY["EPSG","7030"]],
            AUTHORITY["EPSG","6326"]],
        PRIMEM["Greenwich",0],
        UNIT["degree",0.0174532925199433],
        AUTHORITY["EPSG","4326"]],
    PROJECTION["Transverse_Mercator"],
    PARAMETER["latitude_of_origin",0],
    PARAMETER["central_meridian",9],
    PARAMETER["scale_factor",0.9996],
    PARAMETER["false_easting",500000],
    PARAMETER["false_northing",0],
    UNIT["metre",1,
        AUTHORITY["EPSG","9001"]],
    AUTHORITY["EPSG","32632"]]
Origin = (486830.000000000000000,5088540.000000000000000)
Pixel Size = (0.500000000000000,-0.500000000000000)
Metadata:
  AREA_OR_POINT=Area
Image Structure Metadata:
  INTERLEAVE=PIXEL
Corner Coordinates:
Upper Left  (  486830.000, 5088540.000) (  8d49'48.24"E, 45d57'1.09"N)
Lower Left  (  486830.000, 5049400.000) (  8d49'52.08"E, 45d35'52.87"N)
Upper Right (  532370.000, 5088540.000) (  9d25'3.59"E, 45d56'58.80"N)
Lower Right (  532370.000, 5049400.000) (  9d24'54.15"E, 45d35'50.60"N)
Center      (  509600.000, 5068970.000) (  9d 7'24.52"E, 45d46'27.20"N)
Band 1 Block=512x512 Type=Byte, ColorInterp=Red
  Overviews: 45540x39140, 22770x19570, 11385x9785, 5693x4893, 2847x2447
Band 2 Block=512x512 Type=Byte, ColorInterp=Green
  Overviews: 45540x39140, 22770x19570, 11385x9785, 5693x4893, 2847x2447
Band 3 Block=512x512 Type=Byte, ColorInterp=Blue
  Overviews: 45540x39140, 22770x19570, 11385x9785, 5693x4893, 2847x2447
user@hydro1:~/Desktop/LCV_game/como_2010_B$
```

Figure 68: GDAL information for processed Orthophoto.

B. Web Application functionalities

AngularJS directive for the Leaflet JavaScript Library, aims to easily embed maps managed by leaflet on our Leaflet project. Some of the Angular-Leaflet directives which are used in this work have been explained in detail. The figure B-1 shows the functionalities to load a map on the game.

```
$scope.loadTheMap = function () {
    angular.extend($scope, {
        mapView: {
            lat: 45.6,
            lng: 9.9,
            zoom: 8
        },
        layers: {
            baselayers: {
                wms: {
                    name: 'Ortofoto Como 2010 CGR',
                    url: 'http://131.175.143.146/geoserver/CGR/wms',
                    type: 'wms',
                    layerParams: {
                        layers: 'CGR:Como_2010_new',
                        format: 'image/jpeg',
                        attribution: "Orthophoto 2010 CGR"
                    },
                    layerOptions: {
                        maxZoom: 21
                    }
                }
            }
        },
        controls: {
            scale: {
                position: 'bottomright'
            }
        },
```

Figure 69: JavaScript function to load the map on the game

Each time when the player completes one level, another new pixel is loaded for the player. This is made possible by using the Angular Leaflet directives. The figure B-2 shows the functionality, how the GeoJSON layer is updated on the map.

```
$scope.updateTheMap = function () {
        var lat = parseFloat(this.currentPage.lat);
        var lng = parseFloat(this.currentPage.lng);
        $scope.mapView.lat = lat;
        $scope.mapView.lng = lng;
        //$scope.mapView.zoom = 19;
        $scope.mapView.zoom = 18;
        var sLat = lat - 0.000135;
        var nLat = lat + 0.000135;
        var wLong = lng - 0.00019;
        var eLong = lng + 0.00019;
        $scope.geojson.data = {
                    "type": "FeatureCollection",
                    "features": [
                        {
                            "type": "Feature",
                            "properties": {},
                            "geometry": {
                                "type": "Polygon",
                                "coordinates" : [[[wLong, sLat],
                                                [eLong, sLat],
                                                [eLong, nLat],
                                                [wLong, nLat],
                                                [wLong, sLat]]]
                            }
                        }]
                }
};
```

```
        geojson: {
            data: {
                "type": "FeatureCollection",
                "features": [
                    {
                        "type": "Feature",
                        "properties": {},
                        "geometry": {
                            "type": "Polygon",
                            "coordinates" : [[[0.0, 0.0],
                                              [0.0, 0.0],
                                              [0.0, 0.0],
                                              [0.0, 0.0],
                                              [0.0, 0.0]]]
                        }
                    }]
            },
            style: {
                color: 'blue',
                fillOpacity: 0.0
            }
        }
    });
};
```

Figure 70: The functionalities to load new pixels.

There are many operations performed in function.php. Some of the main function has been explained in this chapter. The main operation is to get 5 pixels from the DB to the user for a round. The query for this as follows

```
$query =    "SELECT idResource, pixelLat, pixelLong
            FROM resource
            WHERE idResource NOT IN (SELECT DISTINCT idResource FROM logging WHERE idUser = $idUser)
            AND idResource NOT IN (SELECT DISTINCT idResource FROM resource_has_topic WHERE score > $upperScore)
            ORDER BY idResource LIMIT 5";

$result = $mysqli->query($query) or die($mysqli->error.__LINE__);
```

For each of these extracted pixels, the respective DUSAF and GL30 categories is obtained using the query as follows

```
//GET ALL THE TOPIC RELATED TO THE RESOURCES THAT HAS BEEEN CLASSIFIED BY DUSAF OR GL30
$query =    "SELECT RT.idResource, RT.idTopic, T.value, RT.score,
                CASE
                    WHEN RT.idTopic = R.idTopicDUSAF THEN 'D'
                    WHEN RT.idTopic = R.idTopicGL30 THEN 'G'
                END AS classificationType
             FROM topic AS T
             JOIN resource_has_topic AS RT ON RT.idTopic = T.idTopic
             JOIN resource AS R ON R.idResource = RT.idResource
             WHERE RT.idResource = $idResource
             AND (RT.idTopic = R.idTopicDUSAF OR RT.idTopic = R.idTopicGL30)
             ";

$result = $mysqli->query($query) or die($mysqli->error.__LINE__);
```

In order to fetch the all the categories related to the resource that have been selected by the user we made use of the following query

```
//GET ALL THE TOPIC RELATED TO THE RESOURCES THAT HAVE BEEN EVENTUALLY CHOOSEN BY A USER
$query =    "SELECT RT.idResource, RT.idTopic, T.value, RT.score
             FROM topic AS T
             JOIN resource_has_topic AS RT ON RT.idTopic = T.idTopic
             JOIN resource AS R ON R.idResource = RT.idResource
             WHERE RT.idResource = $idResource
             AND (RT.idTopic <> R.idTopicDUSAF AND RT.idTopic <> R.idTopicGL30)
             ";

$result = $mysqli->query($query) or die($mysqli->error.__LINE__);
```

To fetch all other categories which are neither DUSAF nor GL30 we made use of the following query

```
//GET ALL THE OTHER TOPIC TO FILL ALL THE POSSIBLE CLASSIFICATIONS
$query = "SELECT idTopic, value FROM topic WHERE idTopic NOT IN ( '" . implode($choosenTopics, "', '") . "' )";
$result = $mysqli->query($query) or die($mysqli->error.__LINE__);
```

The final evaluation of the game is performed using the following function within the function.php

```
function evaluation($mysqli, $evaluation){
    $query_evaluation = "select
                IF(SRT.Completed < 1600, SRT.Completed, 1600) AS NofValidatedPixels
                , IF(SL.Played < 1600, SL.Played, 1600) AS NofPlayedPixels
                , SR.Total AS TotalNofPixels
                , SRD.TotalLifePlay AS TotalLifePlayInSeconds
                , CONCAT(FLOOR(HOUR(SEC_TO_TIME(SRD.TotalLifePlay))
                    / 24), ' ', LPAD(MOD(HOUR(SEC_TO_TIME(SRD.TotalLifePlay)), 24),2,0), ':',
                    LPAD(MINUTE(SEC_TO_TIME(SRD.TotalLifePlay)),2,0), ':', LPAD(SECOND(SEC_TO_TIME(SRD.TotalLifePlay)),2,0))
                    AS TotalLifePlayddHHmmss
                , SRD2.NofPlayers
                , SRT2.DUSAF AS NofDUSAFAgreements
                , SRT3.GL30 AS NofGL30Agreements
                , SRT4.NEITHER AS NofDesagreement
                , FORMAT(IF(SRT.Completed < 1600,SRT.Completed/SR.Total*100, 100), 2) AS '%Completion'
                , FORMAT(SRT.Completed/SRD.TotalLifePlay*3600, 2) AS 'Throughput [solved tasks/hour]'
                , FORMAT(SRD.TotalLifePlay/SRD2.NofPlayers/60, 2) AS 'ALP(Average Life Play) [minutes/player]'
                , FORMAT(SRT.Completed/SRD2.NofPlayers, 2) AS 'Expected contribution [solved tasks/player]'
                , FORMAT(SRT2.DUSAF/SRT.Completed*100, 2) AS '%DUSAFAgreements'
                , FORMAT(SRT3.GL30/SRT.Completed*100, 2) AS '%GL30Agreements'
                , FORMAT(SRT4.NEITHER/SRT.Completed*100, 2) AS '%Disagreements'
                from
                (SELECT COUNT(DISTINCT R.idResource) AS Completed FROM resource R JOIN resource_has_topic RT ON R
                    .idResource = RT.idResource WHERE RT.score > 0.94) SRT,
                (SELECT 1600 AS Total) SR,
                (SELECT SUM(TIME_TO_SEC(TIMEDIFF(endRound, startRound))) AS TotalLifePlay FROM round
                    WHERE endRound <> '0000-00-00 00:00:00') SRD,
                (SELECT COUNT(DISTINCT idUser) AS NofPlayers FROM round WHERE endRound <> '0000-00-00 00:00:00') SRD2,
                (SELECT COUNT(R.idResource) AS DUSAF FROM resource R JOIN resource_has_topic RT ON
                    R.idResource = RT.idResource WHERE RT.score > 0.94 AND RT.idTopic = R.idTopicDUSAF) SRT2,
                (SELECT COUNT(R.idResource) AS GL30 FROM resource R JOIN resource_has_topic RT ON
                    R.idResource = RT.idResource WHERE RT.score > 0.94 AND RT.idTopic = R.idTopicGL30) SRT3,
                (SELECT COUNT(R.idResource) AS NEITHER FROM resource R JOIN resource_has_topic RT ON
                    R.idResource = RT.idResource WHERE RT.score > 0.94 AND RT.idTopic <> R.idTopicDUSAF AND RT.idTopic <> R.idTopicGL30) SRT4,
                (SELECT COUNT(DISTINCT L.idResource) AS Played FROM logging L) SL
                    ";

    $result_query_evaluation = $mysqli->query($query_evaluation) or die($mysqli->error.__LINE__);
```

The complete.php is another important file in some of the updating operation has been performed at the end of each level or round. Some of the important operations have been listed here. The user reputation at the end of each round is updated using the query as follows

```
//UPDATE USER REPUTATION

if($levelCounter != 0){

    $updateUserReputation = exp(-($numberOfErrors/2));

    $update_userReputation = $mysqli->query("UPDATE user SET reputation = $updateUserReputation
                                            WHERE idUser = $idUser");

    if($update_userReputation){
        //print $updateUserReputation;
    }else{
        die('Error : ('. $mysqli->errno .') '. $mysqli->error);
    }
}
```

Each round's information, including the timing and score has been updated using the following statement

```
//UPDATE END ROUND AND INSERT SCORE ROUND
$update_endRound = $mysqli->query("UPDATE round SET endRound = '$endRound', score = '$score'
                                   WHERE idRound = $idRound");

if($update_endRound){

}else{
    die('Error : ('. $mysqli->errno .') '. $mysqli->error);
}
```

The next query demonstrates how the player's life play (time consumed for playing) is computed

```
//UPDATE USER LIFE PLAY
$update_row = $mysqli->query("UPDATE user SET life_play = life_play + $lifeTimeInSeconds WHERE idUser = $idUser");

if($update_row){

}else{
    die('Error : ('. $mysqli->errno .') '. $mysqli->error);
}
```

The next set of query demonstrates how the different badges are assigned to the player in the game play. These are designed as switch statements. First is the maximum number of points in the game play is as follows

```
//MAX NUMBER OF POINTS
case 1:
    if ($score == $maxScore){

        //MySqli Insert Query
        $insert_row = $mysqli->query("INSERT IGNORE INTO user_has_badge (idUser, idBadge) VALUES ('$idUser', '$idBadge')");

        if($insert_row){
            if($mysqli->affected_rows > 0){
                $badge['name'] = $name;
                $badge['value'] = $value;
                $badge['image'] = $image;
                array_push($badges, $badge);
            }
        }
        else{
            die('Error : ('. $mysqli->errno .') '. $mysqli->error);
        }
    }
    break;
```

Next badge is all correct answers in the consecutive five rounds.

```
//ALL CORRECT ANSWERS IN AT LEAST 5 CONSECUTIVE ROUNDS OK
case 2:
    $query_allCorrectInConsectiveRounds = "SELECT Q.idUser, Q.score, MAX(Q.count+1) AS maxConsec FROM (
                                    SELECT  @prev_idUser := @idUser AS prev_idUser,
                                            @prev_score := @score AS prev_score,
                                            @idUser := idUser AS idUser,
                                            @score := score AS score,
                                            @count := if(@prev_idUser = @idUser && @prev_score = @score,
                                            @count + 1,
                                            if(@score = 0, 1, 0)) AS count,
                                            @idRound := idRound AS idRound
                                    FROM round ORDER by idUser, idRound
                                ) AS Q
                                GROUP BY Q.idUser, Q.score
                                HAVING Q.idUser = '$idUser' AND Q.score = '$maxScore'";  //AND MAX(Q.count+1)

    //echo "query_allCorrectInAlmost5ConsectiveRounds: ".$query_allCorrectInConsectiveRounds."<br />";

    $result_allCorrectInConsectiveRounds = $mysqli->query($query_allCorrectInConsectiveRounds) or die($mysqli->error.__LINE__);
```

```
// GOING THROUGH THE DATA
if($result_allCorrectInConsectiveRounds->num_rows > 0) {
    //echo "num_rows > 0";
    //echo "result_allCorrectInAlmost5ConsectiveRounds->num_rows: ".$result_allCorrectInConsectiveRounds->num_rows;

    while($row = $result_allCorrectInConsectiveRounds->fetch_assoc()) {
        $maxConsecRounds = $row['maxConsec'];

        if($maxConsecRounds >= 5){
            //MySqli Insert Query
            $insert_row = $mysqli->query("INSERT IGNORE INTO user_has_badge (idUser, idBadge) VALUES ('$idUser', '$idBadge')");

            if($insert_row){
                if($mysqli->affected_rows > 0){
                    $badge['name'] = $name;
                    $badge['value'] = $value;
                    $badge['image'] = $image;
                    array_push($badges, $badge);
                }
            }
            else{
                die('Error : ('. $mysqli->errno .') '. $mysqli->error);
            }
        }
    }
}
break;
```

The next query is for the badge “All the answers in agreement with DUSAF category”

```
//ALL DUSAF ANSWER OK
case 3:
    if ($maxRightAnswerDUSAF == 5){

        //MySqli Insert Query
        $insert_row = $mysqli->query("INSERT IGNORE INTO user_has_badge (idUser, idBadge) VALUES ('$idUser', '$idBadge')");

        if($insert_row){
            if($mysqli->affected_rows > 0){
                $badge['name'] = $name;
                $badge['value'] = $value;
                $badge['image'] = $image;
                array_push($badges, $badge);
            }
        }
        else{
            die('Error : ('. $mysqli->errno .') '. $mysqli->error);
        }
    }
    break;
```

The next query is for the badge "All the answers in agreement with GL30 category"

```
case 4:
    if ($maxRightAnswerGL30 == 5){

        //MySqli Insert Query
        $insert_row = $mysqli->query("INSERT IGNORE INTO user_has_badge (idUser, idBadge) VALUES ('$idUser', '$idBadge')");

        if($insert_row){
            if($mysqli->affected_rows > 0){
                $badge['name'] = $name;
                $badge['value'] = $value;
                $badge['image'] = $image;
                array_push($badges, $badge);
            }
        }
        else{
            die('Error : ('. $mysqli->errno .') '. $mysqli->error);
        }
    }
    break;
```

For all the wrong answers badge the query is structured as follows

```
//ALL WRONG ANSWER OK
case 5:
    if ($numberOfErrors == 5){

        //MySqli Insert Query
        $insert_row = $mysqli->query("INSERT IGNORE INTO user_has_badge (idUser, idBadge) VALUES ('$idUser', '$idBadge')");

        if($insert_row){
            if($mysqli->affected_rows > 0){
                $badge['name'] = $name;
                $badge['value'] = $value;
                $badge['image'] = $image;
                array_push($badges, $badge);
            }
        }
        else{
            die('Error : ('. $mysqli->errno .') '. $mysqli->error);
        }
    }
    break;
```

For the badge "Only three answers before the time ends" the badge is structured as

```
//THREE ANSWER BEFORE TIME ENDS OK
case 6:

    if ($levelCounter == 3){

        //MySqli Insert Query
        $insert_row = $mysqli->query("INSERT IGNORE INTO user_has_badge (idUser, idBadge) VALUES ('$idUser', '$idBadge')");

        if($insert_row){
            if($mysqli->affected_rows > 0){
                $badge['name'] = $name;
                $badge['value'] = $value;
                $badge['image'] = $image;
                array_push($badges, $badge);
            }
        }
        else{
            die('Error : ('. $mysqli->errno .') '. $mysqli->error);
        }
    }
    break;
```

If the player finishes the game within 30 seconds, then the badges assigned is queried as follows

```
//GAME IN LESS THAN 30 SECONDS OR
case 7:
    if ($gameTime != null && $gameTime <= 30){

        //MySqli Insert Query
        $insert_row = $mysqli->query("INSERT IGNORE INTO user_has_badge (idUser, idBadge) VALUES ('$idUser', '$idBadge')");

        if($insert_row){
            if($mysqli->affected_rows > 0){
                $badge['name'] = $name;
                $badge['value'] = $value;
                $badge['image'] = $image;
                array_push($badges, $badge);
            }
        }
        else{
            die('Error : ('. $mysqli->errno .') '. $mysqli->error);
        }
    }
    break;
```

The next query is for the badge “Welcome back after one week of inactiveness”

```
case 8:
        $query_welcomeBack = "SELECT endRound FROM round WHERE idRound <> '$idRound' AND idUser = '$idUser' ORDER BY endRound DESC LIMIT 1";
        $result_welcomeBack = $mysqli->query($query_welcomeBack) or die($mysqli->error.__LINE__);

        // GOING THROUGH THE DATA
            if($result_welcomeBack->num_rows > 0) {

                while($row = $result_welcomeBack->fetch_assoc()) {
                    $endRound_welcomeBack = $row['endRound'];
                    //date in milliseconds
                    $endRound_welcomeBack_milliseconds = strtotime ($endRound_welcomeBack)*1000;
                    // date in milliseconds of 1 week before now
                    $now_milliseconds = strtotime("now")*1000;
                    //one week before
                    $oneWeekBeforeNow = $now_milliseconds - (7 * 24 * 3600 * 1000);
                    if($endRound_welcomeBack_milliseconds < $oneWeekBeforeNow){
                        //MySqli Insert Query
                        $insert_row = $mysqli->query("INSERT IGNORE INTO user_has_badge (idUser, idBadge) VALUES ('$idUser', '$idBadge')");

                        if($insert_row){
                            if($mysqli->affected_rows > 0){
                                $badge['name'] = $name;
                                $badge['value'] = $value;
                                $badge['image'] = $image;
                                array_push($badges, $badge);
                            }
                        }
                        else{
                            die('Error : ('. $mysqli->errno .') '. $mysqli->error);
                        }
                    }
                }
            }
    break;
```

The next query is for the badge “More than 30 min game plays in a single day”

```
        //PLAY MORE THAN 30 MIN IN A DAY OK
        case 10:
            $timeToPlay_30min = 0;
            $query_15min = "SELECT startRound, endRound FROM round WHERE endRound >= CURRENT_DATE AND endRound <  CURRENT_DATE + INTERVAL 1 DAY AND idUser = '$idUser'";
            $result_15min = $mysqli->query($query_15min) or die($mysqli->error.__LINE__);
            // GOING THROUGH THE DATA
                if($result_15min->num_rows > 0) {
                    while($row = $result_15min->fetch_assoc()) {
                        $countGame = $row['startRound'];
                        //date in seconds
                        $timeToPlay_30min = $timeToPlay_30min + (strtotime ($row['endRound']) - strtotime ($row['startRound']));
                    }
                    if($timeToPlay_30min >= 1800){
                        //MySqli Insert Query
                        $insert_row = $mysqli->query("INSERT IGNORE INTO user_has_badge (idUser, idBadge) VALUES ('$idUser', '$idBadge')");
                        if($insert_row){
                            if($mysqli->affected_rows > 0){
                                $badge['name'] = $name;
                                $badge['value'] = $value;
                                $badge['image'] = $image;
                                array_push($badges, $badge);
                            }
                        }
                        else{
                            die('Error : ('. $mysqli->errno .') '. $mysqli->error);
                        }
                    }
                }
            break;
        }
    }
$complete['badges'] = $badges;
}
else
{
    'NO BADGE IN THE APPLICATION';
}
```

The upcoming functions are the most important part of the game shows how the confidence level of the two classifications DUSAF and GL30 is altered according to the selection made by the players.

```
//RESOURCE SCORE
foreach($results as $resultItem){
        $decodedResult = json_decode($resultItem);
        $idTopicSelected = $decodedResult->idTopic;
        $idResourceSelected = $decodedResult->idResource;
        $scoreSelected = $decodedResult->score;
        $userAnswer = $decodedResult->userAnswer;
        $levels = $decodedResult->level;

    switch (true) {

//DUSAF OR GL30 HAS BEEN SELECTED
            case ($userAnswer == 1):
            foreach($levels as $level){
                $idTopicLevel = $level->idTopic;
                $valueLevel = $level->value;
                $scoreLevel = $level->score;
                $resultLevel = $level->result;
                $score_resource_topic_before = scoreResourceTopic($mysqli, $idResourceSelected, $idTopicLevel);

    //IF TOPIC IS THE ONE SELECTED
            if($idTopicLevel == $idTopicSelected){
                //ITS SCORE INCREASES
```

```
			$updateScore = $score_resource_topic_before + ($positiveParameter *
$updateUserReputation);
	//echo "Case 2(selected): Risorsa: ".$idResourceSelected." Topic: ".$idTopicLevel." Before
Score ".$score_resource_topic_before." NewScore ".$updateScore."<br />";
			$update_score_resource_topic = $mysqli->query("UPDATE resource_has_topic SET
score = $updateScore					WHERE idResource = $idResourceSelected
AND idTopic = $idTopicLevel");



							if($update_score_resource_topic){
								//print $updateUserReputation;
							}else{
								die('Error : ('. $mysqli->errno .') '. $mysqli->error);
							}

						}
						//ELSE
						else {
			//IF TOPIC IS THE OTHER ONE
			if($resultLevel) {
				//ITS SCORE DECREASES
			$updateScore = $score_resource_topic_before - ($negativeParameter *
$updateUserReputation);
	//echo "Case 1 TOPIC IS THE OTHER ONE Risorsa: ".$idResourceSelected." Topic:
".$idTopicLevel." Before Score ".$score_resource_topic_before." NewScore ".$updateScore."<br />";
			$update_score_resource_topic = $mysqli->query("UPDATE resource_has_topic SET
score = $updateScore
			 WHERE idResource = $idResourceSelected AND idTopic = $idTopicLevel");
								if($update_score_resource_topic){
									//print $updateUserReputation;
								}else{
									die('Error : ('. $mysqli->errno .') '. $mysqli-
>error);
								}
							}

						}
			}
		 break;
			//NOR DUSAF NOR GL30 HAS BEEN SELECTED
		case ($userAnswer == 0):
			foreach($levels as $level){
				$idTopicLevel = $level->idTopic;
				$valueLevel = $level->value;
				$scoreLevel = $level->score;
				$resultLevel = $level->result;

				$score_resource_topic_before = scoreResourceTopic($mysqli,
$idResourceSelected, $idTopicLevel);
```

```
//IF TOPIC IS THE ONE SELECTED
					if($idTopicLevel == $idTopicSelected){
						//ITS SCORE INCREASES
						$updateScore = $score_resource_topic_before + ($positiveParameter * $updateUserReputation);
					//echo "Case3: Risorsa: ".$idResourceSelected." Topic: ".$idTopicLevel." Before Score ".$score_resource_topic_before." NewScore ".$updateScore."<br />";

						$check_resource_topic = "SELECT idResource, idTopic FROM resource_has_topic WHERE idResource = $idResourceSelected AND idTopic = $idTopicLevel";
						$result = $mysqli->query($check_resource_topic) or die($mysqli->error.__LINE__);
	// GOING THROUGH THE DATA
						if($result->num_rows > 0) {
$update_score_resource_topic = $mysqli->query("UPDATE resource_has_topic SET score = $updateScore WHERE idResource = $idResourceSelected AND idTopic = $idTopicLevel");
								if($update_score_resource_topic){
								//print $updateUserReputation;
							}else{
								die('Error : ('. $mysqli->errno .') '. $mysqli->error);
							}
						}
						else
						{
							$insert_score_resource_topic = $mysqli->query("INSERT INTO resource_has_topic (idResource, idTopic, score) VALUES ($idResourceSelected, $idTopicLevel, $updateScore)");

							if($insert_score_resource_topic){
								//print $updateUserReputation;
							}else{
								die('Error : ('. $mysqli->errno .') '. $mysqli->error);
							}
						}

				}
					//OTHERS REMAIN UNCHANGED
			}
		break;
			} //SWITCH
   }
   echo json_encode($complete);
//END
```