Detection and mapping of illegal settlements in Bhutan using high resolution satellite imagery and cadastral information

By

Pema Zangmo

Thesis
Submitted to Flinders University
for partial fulfilment of the degree of

Master of Geospatial Information Science

College of Science and Engineering
Date: 18/10/2021

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Sl. No	Acronym	Full form
1	MBI	Morphological Building Index
2	OBIA	Object-Based Image Analysis
3	SVM	Support Vector Machines
4	ML	Machine Learning
5	DT	Decision Trees
6	UAV	Unmanned Aerial Vehicle
7	VHR	Very High Resolution
8	GE	Google Earth
9	GSD	Ground Sampling Distance
10	RMSE	Root Mean Square Error
11	DEM	Digital Elevation Model
12	ERDAS	Earth Resources Data Analysis System Imagine
13	AOI	Area Of Interest
14	NN	Nearest Neighbour
15	DEM	Digital Elevation Model
16	RMS	Root Mean Square
17	NIR	Near Infrared
18	GIS	Geospatial Information System
19	CNN	Convolution Neural Network
20	FCN	Fully Convolution Network

21	RPN	Region Proposal Network
22	RoI	Region Of Interest
23	SSD	Single Shot Multi-Box Detector
24	YOLO	You Only Look Once
25	IoU	Intersection or Union
26	TP	True Positive
27	FP	False Positive
28	FN	False Negative
29	TN	True Negative
30	AP	Average Precision
31	mAP	Mean Average Precision
32	VOC	Visual Object class

ABSTRACT

The rampant growth of illegal settlements is drawing attention worldwide and is becoming one of a most concerning issue for the Governments all over the world. Different countries have introduced many preventive measures and attempted to understand the very core rationale for the growth of illegal settlement by way of introducing (for e.g., Land pooling concept in several cities and affordable housing scheme and resettlement programs) but some countries have instigated restrictive measures to curb such practices with laws and regulations in place. The general observation for such settlement is that they arise from rapid urbanization and lower rungs of society facing the brunt of high land prices, lack of affordable housing, expensive cost of living and economic disadvantages. The nature of propagation of illegal settlements remains similar worldwide, but the approach to mitigate the issues remains ambiguous. One of the issues that Governments have to contend with is the location of such illegal settlements. Remote sensing, particularly using satellite imagery, for detection and mapping have been gaining attention and are being adopted worldwide for its efficiency and optimal resource allocation.

This project is carried out in a southern district of Bhutan and the main aim for the project is to semi-automatically detect and map buildings using very high-resolution imagery (GeoEye-1 and Pleaides-1A imagery) and then delineate illegal buildings using the legal parcels as documented in cadastral land information. There is a slight deviation in the upsurge of illegal settlements in Bhutan when compared toother countries; issues regarding the long-standing land tenureship and parcel boundary anomalies can be the most probable factors that have triggered the growth of illegal settlements. Bhutan is still a developing country; hence the accessibility and coverage of road and other infrastructure are still in the process of development. The nature of illegal settlements is not as rampant when compared to other countries worldwide but there were few cases that were reported in the recent newspapers which became very concerning to the authorities. The Land department has then taken up the responsibility to put a system in place to restrain the impending concerns and the measure includes employing of land sector inspectors in every district and block for scrutinization and reporting the illegal building cases; this is slow and expensive. No research has been carried out in Bhutan using remote sensing capabilities and imaging techniques for mapping which potentially provides an optimal solution for the detection.

GeoEye-1 imagery acquired by the Bhutan government in 2013 was used as primary data source for the project and the methods for detection and mapping includes: 1) image pre-processing and data preparation, 2) Pan sharpening using suitable methods, 3) image segmentation and classification using object-based image analysis 4) preparation of primary data for accuracy assessment, 5) accuracy assessment of the classified image, 6) delineation of illegal settlements. When conducting the preliminary data assessment of the GeoEye-1 imagery it was discovered that both Panchromatic (PAN) and Multispectral imagery (MS) had significant co-registration issues and

required orthorectification using numerous control points. The pan sharpening method produced poor results as the co-registration of PAN and Multispectral imagery was not accurate. Further the area of interest was limited to a smaller section from the imagery owing to the topographical attributes and undulating terrains in Bhutan, which resulted in anomalies when the orthorectification was carried out over a larger area. Hence the image segmentation and classification were restricted to GeoEye-1 Multispectral imagery of resolution 2m and the processes involved were conducted using in ArcGIS Pro software. The data to be used for accuracy assessment contained point and polygon building data collected by the Bhutan Government using GPS and other surveying instruments in the year 2014; but the data was not exhaustive and needed visual inspection and manual digitization for some parts of the areas keeping the base reference as the GeoEye-1 Multispectral imagery data for year 2013. The final classified image using Object Based Image Analysis (OBIA) was assessed based on the improvised building data. All geospatial analysis were done in ERDAS Imagine 2018 and ArcGIS Pro version 2.8.0. Object based image classification resulted in detection percentage of 99.5% and the overall correctness of the model was 51.5%. In line with the previous literatures, the spectral similarities in the road surface and building rooftops lead to misclassification and affected the total accuracy.

Another high-resolution imagery Pleiades-1A was acquired for time period 5/11/2017 on 09/10/2021 with the scene area of 80sq.km. Another approach for detection of buildings using deep learning algorithm was investigated. The method used for deep learning includes: 1) Image pre-processing, 2) Pan sharpening using suitable methods, 3) selection of suitable model for object detection, 4) Accuracy assessment, 5) Delineation of Illegal settlements. The image processing and analysis were conducted using ERDAS Imagine 2018 and ArcGIS Pro version 2.8.0 softwares. The detection precision was 20% for IoU (Intersection or Union) less than or equal to 10%. The detection model produced a poorer accuracy and there were numerous underlying issues that contributed to the such discrepancies which are discussed in the discussion chapter of the research.

The major findings from this project are not limited to mapping the building footprints but, with the buildings feature output as produced using the OBIA methods and deep learning will be overlaid with the cadastral information as provided by the government of Bhutan to identify and map illegal settlements. The basic thumb rule for identification of illegal settlements would be that building that falls outside of the legal parcels would be considered illegal in that respect. Many a times the term “informal” is synonymously used with illegal and there is no clear dichotomy in defining the informal by compromising the illegality aspect. Especially in the case of Bhutan the rise of informal settlement is mostly in line with the legality issues rather than the physical condition/attributes defining a settlement.

Keywords: *Illegal, Informal, Settlement, GeoEye-1, Pleiades-1A, OBIA, Deep learning, Orthorectification.*

DECLARATION

I, Pema Zangmo, certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Signed: 

Date: 18/10/2021

ACKNOWLEDGEMENTS

Thesis is not just an individual endeavour, it's a journey you have to go through but thankfully I had the privilege of such inspiring and enthusiastic team to whom I now would like to pay my utmost respect for making the journey a worthwhile.

Firstly, I would like to extend my heartiest gratitude to the Flinders University and the course coordinator for providing us with the opportunity to undertake an extensive study and shaping our budding potential which was thus disclosed by in by.

I would also like to thank the most important pillar to my thesis, Professor David Bruce of Flinders University who guided me all along. A very rudimentary idea for my research was patiently and precisely moulded with his technical expertise and unwavering support. He consistently allowed this thesis to be my own work and steered me in the right direction whenever required. Without the unwavering support and dedication, it wouldn't have been a swift journey and I owe him my utmost gratitude and respect.

I would also like to express my gratitude to Professor Robert Keane, Professor Tessa Lane, School of the Environment, Flinders University, for grooming us and facilitating us with constant support and assistance.

My sincere gratitude and appreciation to my peers and Mr. Dorji Tashi for being my emotional support system. The educative conversation and suggestions really helped me get a better perspective to my thesis.

I would like to extend my appreciation to Mr. Tenzin Namgay, Director, Mrs. Sonam Yangdon, Mr. Kinga Loday, Deputy Survey Engineer and Mr. Pembu Tshering, National Land Commission, for facilitating and entrusting me with the primary data which was indeed the backbone to my research.

I would like to thank my parents for their support and encouragement throughout my research work. Finally, I would like to thank all the persons whose names have not been mentioned above but helped me directly or indirectly in their own ways for the completion of my thesis.

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

1 CHAPTER: INTRODUCTION

1.1 Problem statement/Background

Numerous endeavours have been invested in defining factors and characteristics of informal settlements that confines from a formal settlement (Sietchiping, 2005).Informal settlement has multitude of defining characteristics that labels an informal settlement and each countries defines characteristics of informal settlement differently Okongo (2019), (Hasan, 2016). Generally informal settlement in various countries encompasses the physical attributes and living standards of people residing (Hasan, 2016, Ojwang, 2009) but informal settlement in Bhutan is moreover aligned to the illegality of the land tenureship. The term “illegal settlement” is often defined in conjecture with the informal settlement and slum like conditions hence defining illegal settlement embodies a lot inconsistency and vagueness (Gram-Hansen et al., 2019).

Growth of informal settlements are a global concern and with accelerating population growth, rapid urbanization, lack of governing policies and implementation of affordable housing schemes have contributed to rampant growth of informal settlements (Samper et al., 2020). Around 25% of world’s urban population reside in informal settlements and since 1990 there has been an addition of around 213 million residents in informal settlements (Avis, 2016). In many countries as a preventive measures various improvement and slum upgradation programmes by way of *in situ* upgradation or relocation and rehabilitation to a formal housing are implemented (Li et al., 2005, Hasan, 2016, Thomson et al., 2020).But in few countries it fails to accept the existence of growing informal settlements by placing strict rules like eviction undermining the embedded link to local economy and social network (Avis, 2016, Arroyo, 2011). Various approaches which include field data collection, using UAV and photogrammetry to map the informal settlement have been opted but for monitoring and effective management purposes, use of geospatial and remote sensing capabilities are recommended as a cost and time effective endeavour in most countries (Li et al., 2005, Arroyo, 2011 (Thomson et al., 2020)).

Bhutan is also no exception to such ramifications as a result of rapid urbanization and population growth. An increase in 30% of urban population was reported during 2005 census from 15% during 1999 census enumeration (MoWHS, 2016). The concept of informal settlement in Bhutan is encircled around the legality of the plots rather than the generalised concept of an informal settlements as prevalent in other countries. In the last few years, it was reportedly covered in the national newspaper and news bulletin board regarding the increasing illegal settlement in the southern belts of Bhutan which shares a close border to the

neighbouring country like India. Such rise of illegal settlements has become very concerning for the authorities and hence it needs to foresee a proper system in place to avoid the upheaval which are prevalent in other countries.

Various adoption of Remote sensing techniques that includes object and pixel based image classification (Marangoz et al., 2006, Fallatah, 2020), spatio temporal modelling (Dubovyk et al., 2011), statistical and regression modelling techniques (Gavankar and Ghosh, 2018) , Morphological and shadow index modelling (Xin and Liangpei, 2012, Benarchid et al., 2013), using UAV technology (Gevaert et al., 2017) and classification infused with machine learning s (Gram-Hansen et al., 2019, Norman et al., 2021) have been adopted in detecting and mapping informal settlements.

There are both benefits and limitations to the adoption of various technologies and approaches available. Satellite imaging technology and products are considered as the most reliable source for such research owing to the spatial and spectral performance but the limitations associated with classification of the overlapping spectral similarities and shadow detection have been a challenging task (Blaschke, 2010, Pushparaj and Hegde, 2017). Exploration in use of LiDAR data using UAV technology have successfully produced high quality images over smaller areas but it is an expensive and tedious endeavour to carry out over large areas (Gevaert et al., 2017). However, use of machine learning and deep learning algorithms have been gaining a lot of attention for image classification, object detection as well as change detection studies; the approach has the capacity to handle large datasets and is a cost-effective method (Madhavan, 2017, Norman et al., 2021, Gram-Hansen et al., 2019).

1.2 Research Aim and objectives

The main aim of the research is to detect and map illegal settlements in study area in Bhutan using very high-resolution satellite imagery and cadastral information.

The objectives of the study area are to:

- Critically select suitable satellite imagery and the approach to classify or detect building footprint.
- Examine the accuracy of the output.
- Delineate illegal settlements using cadastral information
- Provide suggestions and recommendations for the output produced.

1.3 Introduction to Bhutan

a) Location and Topography

Bhutan is a small land locked country sandwiched between two of the most populated countries namely China in the north and India in the south (Bureau, 2020). The total area of Bhutan is 38,394 square kilometers with east west stretching 300Km and north south 170 Km. Bhutan is a mountainous country with about 70% of total country's area covered in forest (MoWHS, 2016). Bhutan is landscaped with rugged mountains and undulating terrains. The elevation ranges from 160 meters to more than 7500 meters above sea level. The entire country is divided into three altitudinal regions namely Himalayan zone (above 4500m altitude) areas significantly encompasses snowcapped mountains and above tree line, temperate zone (1000-4500m) and subtropical region characterized with southern foothills below 1000m altitude and valleys below 500m altitude (Bureau, 2020).

b) Temperature and precipitation

The temperature varies according to the relief and mostly the central part of Bhutan experiences a cool and temperate climate while in southern Bhutan is it rather hot and humid during summer with temperature ranging from 15-30 ^{0}C. In the northern part of Bhutan the precipitation is mostly in the form of snow, the central Bhutan experiences an annual rainfall of around 1000mm and the southern Bhutan has average precipitation of 7800mm (Bhutan, 2017).

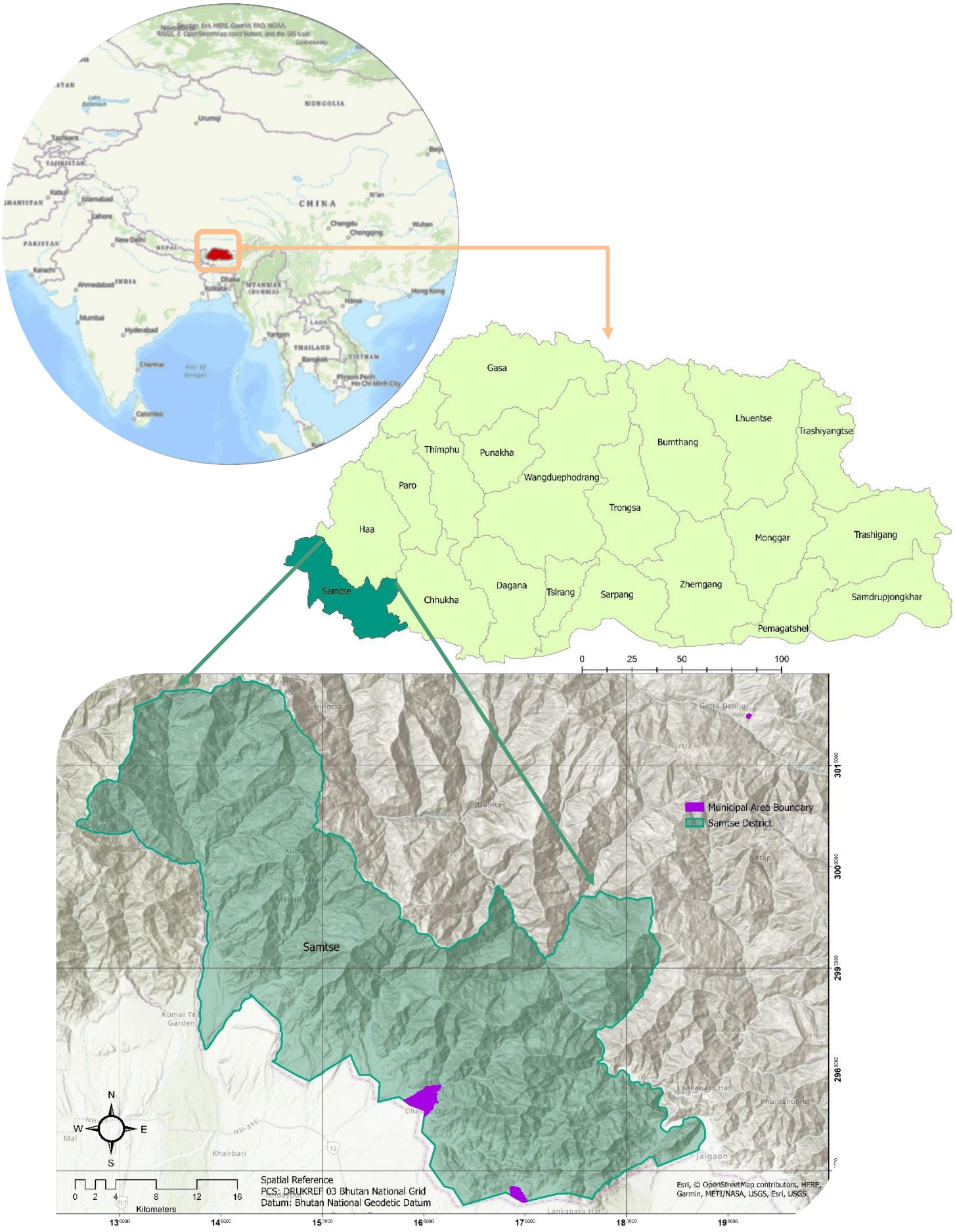
c) Population and demographic profile

Bhutan has enumerated population and housing data for only two census years namely in 2005 and the second one in 2017. In 2005 the total population recorded was 634,982 persons and male and female population was 333,595 and 301,387 respectively (MoWHS, 2016). In 2017 the total population enumerated was 735,553. The total urban population has increased from 196,111 persons in 2005 while in 2017 it is recorded as 274,316 persons (Bureau, 2020).

1.3.1 Introduction to Region of interest (Study area)

Samtse is located on the southern part of Bhutan sharing a boundary with neighboring country India and the district accommodates 62,700 population. The total area of the district is 1,305Sq.km and the elevation profile of the district is around 200-4400m above sea level (Bureau, 2020). The district consists of large-scale cement, steel and quarry industries and many other medium industries. The region has two sub class town areas as depicted in Figure 1.

Figure 1. Location map for Bhutan and Samtse district (Region of interest)



1.4 Research Questions

a) What is the minimum size of the dwelling units in the area of interest?
b) What is the minimum spatial and spectral resolution required to identify typical building structures in Bhutan?
c) What are the likely environments to identify informal settlements in Bhutan?
d) What are the most appropriate methods for detecting buildings in high resolution satellite imagery?
e) What are the most appropriate techniques for accuracy assessment of the data generated using various methods?

1.5 Potential significance of the research

The project will be a pioneer endeavour in detection and mapping of illegal settlements using remote sensing and GIS capabilities. Until this point in time, there has been little effort or attempt to use such technology in addressing the prevailing issues. Increasing illegal settlement has become very concerning to the government and in response to that, government took a swift action by employing land inspectors in all the 20 districts in Bhutan to report on the illegal settlements by validating the cases on ground and then consulting the land department for further directives. The current approach definitely has its benefits as a part of ground truthing and interaction with the affected households on the genuineness of the case. But such approach is very time consuming and demands a lot of human and financial resources. Relying on one official to cover an entire district and report on the issues would be a very tedious and slow process but moreover the high probability of unreliable system as the data captured can be subjected to biasness and human errors. This project can indeed be a pivotal moment in approaching the issues from remote sensing and GIS perspectives. Incorporation of the modern technologies such as geospatial science, can be a viable option as it is both time and cost effective and larger area of interest can be addressed.

“Landlessness” is also one of the concerns in Bhutan and to alleviate the living standards of the people, His Majesty, the king of Bhutan, initiated the new process of granting lands to the landless and socio-economically disadvantaged population, considering all the ground situations and eligibility. The rehabilitation and resettlement program was initiated in the year 1997 in five districts benefiting 4,600 households (NLC, 2012). In this program, the district and block level officials report to the regional coordinator for eligible households for the land grants. The coordinator corresponds with head office for the further proceedings of the case. Despite the proactiveness of the system, it fails to incorporate modern technologies, such as geospatial science (NLC, 2012).

This particular research which explores various semi-automated models/approaches using satellites imageries to identify the vulnerable section of households can be a major input for the program and the utility of such advanced technology can ensure productivity of the program and wider outreach of people.

1.6 Outline of the research

CHAPTER 1. Introduction	This chapter present the rationale for carrying out the research, statement of aim and objectives, the research questions section that directs the orientation of the research, and then the significance of the project.
CHAPTER 2 Literature review	This chapter commences with a global context on the informal settlement and then defining the informal settlement and slum. The literature review on the mapping and management of informal settlement globally that doesn't include the use of remote sensing technologies. Then a detailed review of detecting and mapping informal settlement using remote sensing capabilities and other approaches are presented. An overview on the pan sharpening methods, object-based image analysis and deep learning approaches is provided. The chapter present an overall methodology for the research.
CHAPTER 3 Methods	This chapter presents detailed methodology that includes details on data acquisition, image pre-processing, two classification methods namely: Image classification using Object based Image Analysis for GeoEye-1 Imagery and deep learning approach for Pleaides-1A Imagery and accuracy assessment using both approaches and finally delineation of illegal settlement using both imageries.
CHAPTER 4 Results	This Chapter presents the results from image pre-processing, image classification, Object detection and accuracy assessment for both approaches.

CHAPTER 5 ***Discussion***	Discussion on the results from the approaches adopted.
CHAPTER 6 ***Conclusion and recommendations***	Conclusion and recommendations, statement of limitation of the projects and way forward for the project

Table 1. Outline of the research project

1.7 Overall methodology

Figure 2. illustrates the overall methodology adopted for the research starting from the initial thought process to the data collection, image processing and the conduction of satellite imagery accuracy assessment for the selected approach and finally the delineation of illegal settlements.

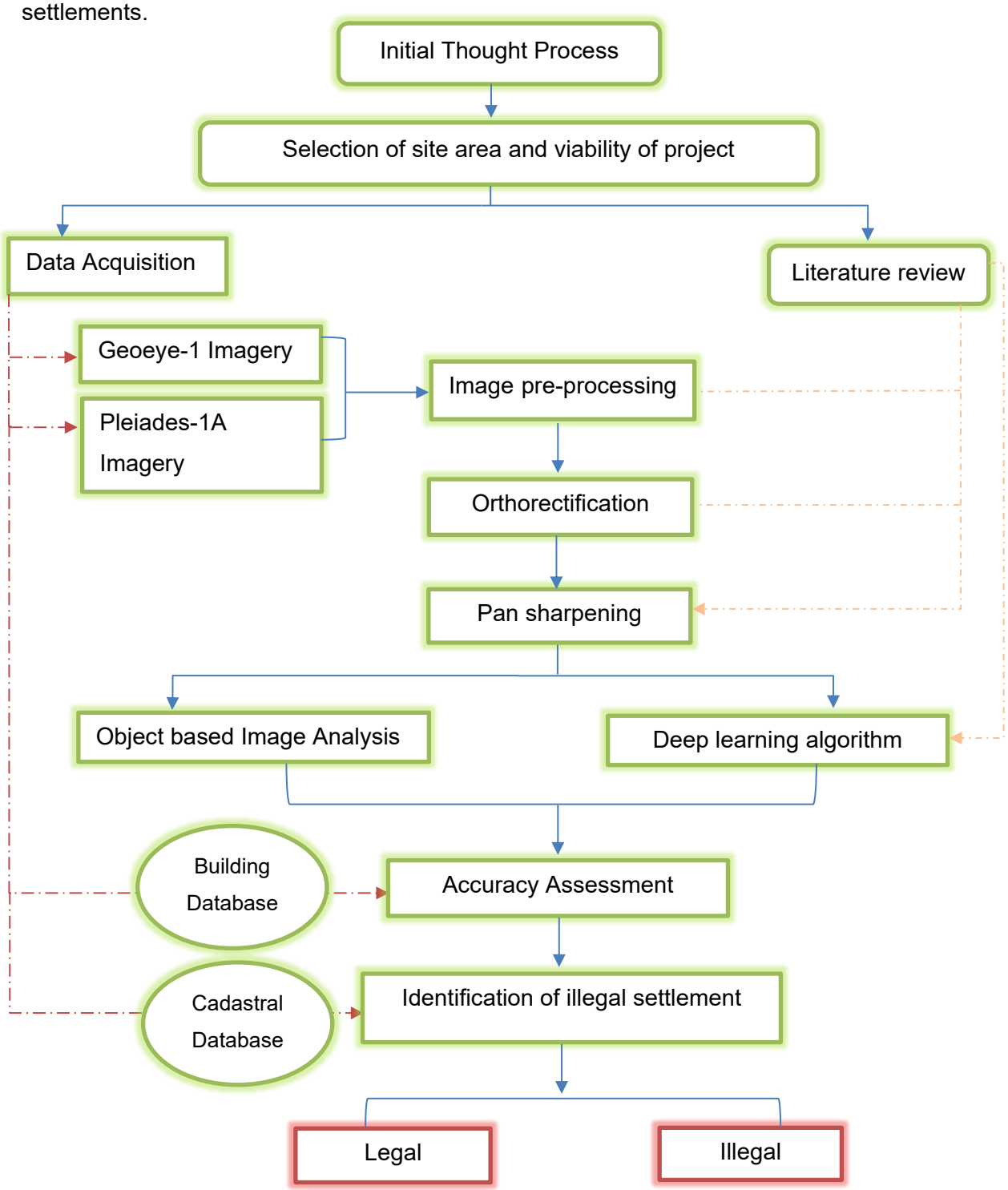


Figure 2. Overall methodology for the research

2 CHAPTER: LITERATURE REVIEW

2.1 Global context on informal settlement

Rampant growth of slums and informal settlement is reportedly becoming a global concern. According to the United Nation sources, it is reported that over 50% of the world's population is residing in slum like conditions or informal settlement. The growth of informal settlement is comparatively rapid in developing nations (Graesser et al., 2012). Rapid urbanization of the cities has led to huge consumption of the scarce land resources and caused increasing urban sprawl around the cities (UN-Habitat, 2020). The vertical and horizontal expansion of city spaces are evidently growing faster as a response to the increasing urban population and consequently exerts enormous pressure on the limited land resources (UN-Habitat, 2020). One of the major driving forces of rapid urbanization is the migration, and according to the report (UN-Habitat, 2020) , there are 763 million internal migrants and 272 million international migrants in the world currently. Migration accounts to a lot of spatial transformation in terms of meeting the demands for better infrastructures and basic amenities for better livelihood but also brings diversities and heterogenous cultural and ethnic fusion which eventually carves the city spaces (UN-Habitat, 2020). With the increasing number of people living in the urban areas, it leads to an issue of affordable housing around city centre (Hasan, 2016). High land prices around the city and availability of facilities uplift the living standard of city areas which consequently impacts the low-income groups (Avis, 2016). Globally it is reported that an household spends approximately 25% of the income on the rent (UN-Habitat, 2020). To respond to the increasing demand for affordable housing, the city keeps expanding and areas around city are further consumed for development to meet the needs of the people (Thomson et al., 2020). Globally such issues are the major cause in the rise of informal settlement and slums in the regions (Dovey et al., 2020).

2.1.1 Definition of Slum

The slum like condition was first identified sometime around in 1820s and it was basically used to refer to poor quality of housing and unsanitary living conditions of people where the outbreak of violence and drug abuse was more rampant (UN-Habitat, 2007). Traditionally, a slum was inhabited in a liveable neighbourhood which eventually upon progressive ownership transfer and subdivisions lead to deterioration of the quality of structures and unlawful constructions (UN-Habitat, 2007). In short, a densely populated urban areas with substandard infrastructures and services were defined as slum (Graesser et al., 2012). By in by, the definition of the term "slum" altered with the changing environment of the world (UN-Habitat, 2007). Today, Slum includes informal settlements with structures ranging from temporary sheds to well-maintained buildings but generally characterised by lack of basic facilities,

sanitations , land tenureship discrepancies and below poverty line groups (UN-Habitat, 2007, Graesser et al., 2012).

UN Expert group defines “slum” as having irregularity, informality and illegality aspect to urban morphology (Dovey et al., 2020, Graesser et al., 2012):

1. No proper access to safe drinking water,
2. Inadequate sanitation and other services,
3. Dilapidated and old housing structures,
4. Overcrowding (more than 3 person per room (Dovey et al., 2020) with no green space/open space,
5. Illegal land tenureship and no sense of security (Graesser et al., 2012)

2.1.1.1 Factors contributing to the Existence of slums

1. **Rural Urban Migration:** Since 1950, in the developing countries, the people engaged in agricultural sectors decreased from 20 to 30% and the rural-urban migration drastically increased (Avis, 2016). Many people migrated to the urban areas for employment and better opportunities and in many cities informal sectors account to 60% of the urban population employment (UN-Habitat, 2020, Graesser et al., 2012). Increasing growth of informal sector has led to rapid growth of slums and poverty (UN-Habitat, 2007). According to UN-HABITAT projection by the year 2030, in Africa most population will reside in cities/urban areas and rural area will cease to exist. Latin America has also undergone rapid urbanization and currently it accommodates 75% of population in urban areas (UN-Habitat, 2020). Asia is home to half of the world population and as a result of rapid urbanization, 36% of the population lives in cities. Some of the largest cities in Asia like Mumbai , Calcutta and Bangkok have population exceeding 10 million people and out of which almost one half of the people live in slum conditions (UN-Habitat, 2007). In Nairobi, Kenya 60% of the population resides in slums and squatter settlement in 5% land area with high density structures and inadequate amenities (Odongo, 2017).
2. **Discrepancies in land tenureship:** Many slum dwellers reside in areas without secure land ownerships and without the entitlement**,** the area devoid of amenities and formal stationing of services and infrastructures leading to more slum like conditions (UN-Habitat, 2007, Graesser et al., 2012).
3. **Affordable housing :** There is a direct relationship in the growth of informal settlements and slums with the availability of adequate and affordable housing suiting all rungs of income slabs (Graesser et al., 2012)
4. **Globalization:** The requirement of skilled labours, inequality in wealth distribution, global economic boom have brought negative implications in the growth of slums.

There is a very fine line in defining informal settlement from slum. A Slum is rather a condition/environment the people in the informal settlement are experiencing because of the poverty, dilapidated housing conditions , no proper access to basic facilities and sanitation in the areas they are inhabiting (III, 2015).

2.1.2 Defining informal settlement

An informal settlement is characterised as a residential area (III, 2015, Avis, 2016, Graesser et al., 2012, Gram-Hansen et al., 2019) with:

1. No legal land or dwelling tenureship,
2. Neighbourhood lacks basic infrastructure and services like road, water and power supply,
3. Structures and planning not in compliance with the building rules and regulations.
4. The modality of inhabiting ranges from squatter settlement in the area to illegal rental housing.
5. Often situated in environmentally/geographically sensitive areas.

Some literatures define informal settlement as a settlement which are unplanned and has irregular structures, varying use of building materials such as plastics, wood and tin sheets, densely populated area and is an immediate product of increasing demand of housing for urban poor (Li et al., 2005)

Existence of slums and informal settlement in the cities areas does affect the overall city outlook and sustainability for which many countries adopt the forced evictions but there are countries which appreciates its existence as consequential phenomenon brought about by city space transformation and developments. They acknowledge the mixed land use of informal sector in the city plan and the wide ranges of employment it provides to the urban dweller despite the deviation from main stream urban opportunities (DDA 2007). The informal sector often supports and compliments the city environment by means of low wage laborers and a part of subsistence economy (III, 2015).

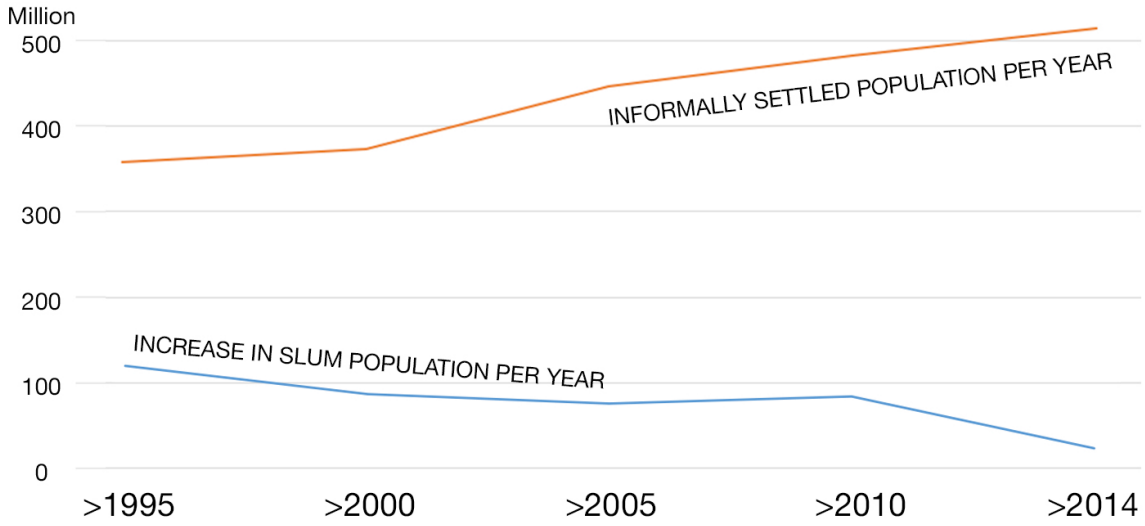


Figure 3.Trend of Slum and informal settlement from 1995 to 2014 (Dovey et al., 2020)

The 'informality' is still a term with numerous ambiguities and often intertwined with slums conditions and the meaning/perspectives adoption differ in various countries (Dovey et al., 2020, Gram-Hansen et al., 2019). In various developing countries urban population have been accommodated and informally settled as a part of government policy and plan implementation. Over the decades the interventions in addressing the increasing slums and informal settlement issues has considerably lead to decrease in the people residing in slums as depicted in Figure 3 (Dovey et al., 2020). The differences between slum and informal settlement cannot be concluded nor the synonymity can be well defined as level of adversity/ripple effect of such widespread cannot be comprehended either (Dovey et al., 2020).

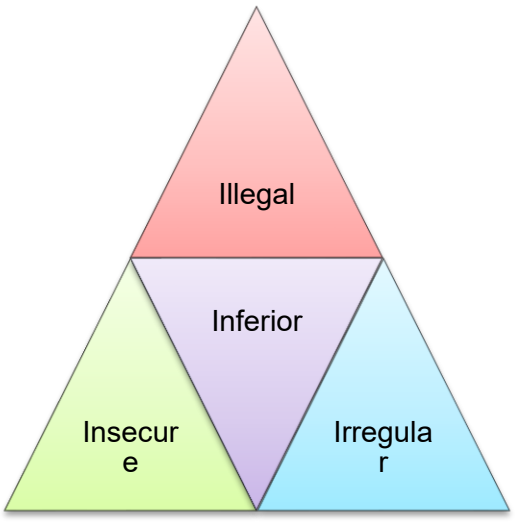


Figure 4. The four i's of urban informality (Dovey et al., 2020)

It can be concluded that informality exists beyond the urban poor – the slum is one kind of informal settlement that is defined as illegal or inferior and rendered insecure by being targeted for eviction/demolition (Graesser et al., 2012). Nevertheless the status quo of rampant growth of informal settlement/illegal settlement reflects on the causal effect of rapid urbanization and overall living environment of cities. Therefore, a reliable and efficient spatial intervention is essential for upgradation of living conditions and assist the decision-making process for any actions (Blaschke et al., 2008). Use of high resolution satellite imagery and geospatial analysis can facilitate in providing relevant information for the parametrization of urban landscape based on physical properties (Graesser et al., 2012). With the power of high-resolution imagery, the delineation of informal settlement is possible owing to the distinct physical and contextual properties of slums and informal settlement (Gram-Hansen et al., 2019).

2.2 Various approaches adopted to address/monitor informal settlement all over the world

This section presents a review on approaches adopted by various countries which doesn't completely incorporate the use of remote sensing capacities for mapping informal settlement but the imageries perform as a backbone for carrying out other surveys.

In Kibera, Kenya the mapping of informal settlement was conducted by empowering the citizen, communities and local government by leveraging mobile phones and GIS applications for collecting location of informal settlement and location data of the level of services available in the area. It was a participatory approach as a part of digital mapping and ground truth mission carried out in 2010. The pilot project in one slums was successful and shifted into second phase in other slums following similar approach and it was further strengthened with support and coordination from respective authorities to empower the capabilities of using maps for such purposes Hagen (2011).

Extensive research was carried out in Kibera which adopted the use of Satellite imagery (Quickbird imagery acquired in 2004) to facilitate the identification of infrastructures available on ground and Landsat Thematic Mapper (TM) imageries for years 1976, 1987, 1995 and 2002 was analysed for temporal land use changes specifically to detect urban sprawl and expansion (Ojwang, 2009). Furthermore, survey questionnaires and infrastructure mapping were conducted using handheld GPS in the area of interest to further validate the urban sprawl. The existing condition(Quality of life) in the area with respect to modes of water supply, availability of public toilets and sanitation, street infrastructures, drainage and sewerage system, Solid waste management, Migration pattern and mode of transportation were also analyzed to compliment the time series analysis (Ojwang, 2009) .

Another study was conducted but the approach was not only limited to the mapping of informal settlement but also to understand the spatial pattern and influencing factors contributing to such accelerated growth. In Istanbul, Turkey there were numerous studies that focused on mapping informal settlement and growth per se. The study (Dubovyk et al., 2011) incorporated the available and well-established data like land use maps for different time periods(1990, 1995, 2000, and 2005), Digital elevation models and informal settlement map for year 2006, environmental and transport data, population and demographic data. In this research the method adopted was the Logistic regression modelling and evaluation. Various factors maps and binary maps of informal settlement was produced for each time periods to generate the trends and understand the relationship. Check of multicollinearity was conducted to determine the factors and select the model for the studies. The fitted curve model was then used for calculation and Likelihood of informal settlement growth. The result from the research were assessed and validated by comparing to similar studies done in past and the Logistic Regression model inheriting spatial explicitness is thus considered to have an upper hand when compared to other models. The model showed its relevancy and consistency when compared to the ground reality and can be a useful tool in facilitating urban planning processes (Dubovyk et al., 2011).

A lot of studies are delimited with the lack of basic and exhaustive geodata to identify the informal settlement hence the use of satellite imagery is particularly helpful in extracting the physical attributes of informal settlements. Informal settlement is generally characterised with overcrowding, irregular structures and narrow roadways; the resolution of satellite imagery available doesn't suffice the requirement to map such high detailed features/region (Thomson et al., 2020). The use of photogrammetry and mapping using UAV is rather a cost effective and flexible solution in acquiring high resolution imagery but the drawback in using such technology is the spatial extent for flights and the processing factor which rules out the use of such technology for larger area of interest. A study was carried out using UAV and the output processed was DSM and Orthomosaic of resolution 3cm for two study sites (Gevaert et al., 2017). The point cloud was then analysed using 10 different classes and the Support Vector Machine (SVM) Classifier was used to perform the classification to identify the object and then allocate the attributes related to the objects. The author suggested the use of 3D point features to be included in the processing (Niemeyer et al., 2014) along with the other datasets like mean shift segmentations and DSM that resulted in an accuracy of 86% (Gevaert et al., 2017)

2.3 Pan sharpening of raster data

Pan Sharpening is a spatial transformation which combines images of different resolutions to increase the spatial quality and spectral enhancement of the fused image when compared to the original Multispectral imagery (Jawak and Luis, 2013). It is an essential processing that prepares the raw imagery for further analysis and classification. There are numerous pan sharpening method/algorithms available (Zhang and Mishra, 2012) and fusion of Multispectral and Panchromatic imageries enables the further treatment to the images for spectral and spatial resolution enhancement (Sun et al., 2014). Sinch the launch of SPOT-1 satellite in 1986 which acquired Panchromatic image with resolution 10m and Multispectral image with resolution 20m, it initiated the exploration and researches in pan-sharpening the imageries (Zhang and Mishra, 2012). With the advanced and modern imaging sensors, it has enabled the delivery of high resolution imagery in terms of Multispectral as well as Panchromatic images (Sun et al., 2014). Trade off in terms of spectral and spatial resolution is still very evident and the applicability is dependent on the resolution of imageries (Sun et al., 2014).

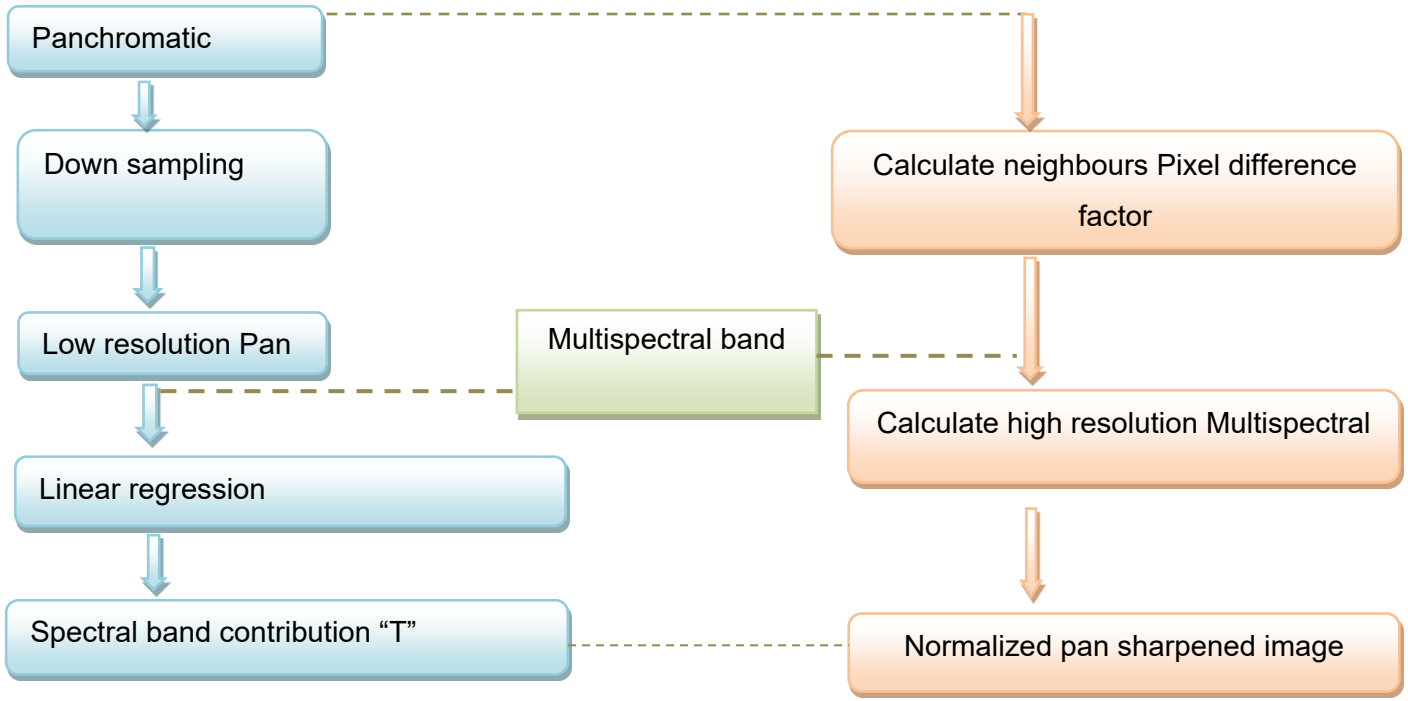
Table 2. Various Pan Sharpening methods available

Sl. No	Pan sharpening Method	Specification
1	**High pass filter (HPF) Resolution Merge**	This method extracts high frequency information from the Panchromatic imageries using high pass filter (HPF). The extracted high frequency information is then added to the Multispectral imageries with specified weights (Zhang and Mishra, 2012). The method calculates ratio “R” which depicts the spatial size of the Panchromatic and Multispectral images and the high pass filter is added on the Panchromatic imageries. Resampling of pixel size for Multispectral images is done to the High pass Panchromatic images (Pushparaj and Hegde, 2017). Then the merging of high pass filter Panchromatic and resampled Multispectral is conducted for high spatial and spectral resolution enhancement (Pushparaj and Hegde, 2017) .

2	**Resolution Merge (Principal Components)**	This method firstly converts the original Multispectral image into an uncorrelated new set called principal components (Pushparaj and Hegde, 2017). The first principal components usually correlate to the Panchromatic image and is considered a good indicator of spatial information whereas the succeeding principal components relates to the spectral data of Multispectral image (Pushparaj and Hegde, 2017) . Evaluation of the histogram of the Panchromatic with the first principal component must be matched before the inverse transformation to avoid the color distortion (Zhang and Mishra, 2012). First principal component is then replaced by the high resolution Panchromatic image and then an inverse transformation is conducted to generate merged image (Zhang and Mishra, 2012). The basic assumption for the fusion relates to the overlap of Multispectral bands with the high resolution imagery and the quality is strongly dependent on the degree of spectral overlap between the two imageries (Zhang and Mishra, 2012).
3	**Gram–Schmidt Spectral sharpening method**	By averaging the Multispectral bands, a lower spatial resolution pan image is replicated. To the replicated Panchromatic bands and low-resolution Multispectral bands, Gram Schmidt transformation is applied (Zhang and Mishra, 2012). The replicated Panchromatic bands are considered as the first band and thus high-resolution pan image is adjusted to comply with the transformed bands (Pushparaj and Hegde, 2017). The adjusted high resolution Panchromatic is thus used to replace the first band of the transformed bands and inverse Gram Schmidt inverse transformation is applied to generate the pan sharpened Multispectral bands (Pushparaj and Hegde, 2017). The procedure resembles the principal component method (Zhang and Mishra, 2012). Generalized minimum mean square-error (gMMSE) method is an optimised and multivariate statistics-based approach that uses general Laplacian pyramid (Sun et al., 2014). In this method, high frequency resolution of Panchromatic imagery is injected to the Multispectral imagery and the methods involves minimising the root mean square errors by optimising the parameters involved in the fusion (Sun et al., 2014).

4	**The Brovey transform method**	It is a simple and classical fusion methods that uses an algebraic expression to inject the overall brightness of the Panchromatic image into every pixel of Multispectral image (Sun et al., 2014) and enhances the visual appearance of the image (Pushparaj and Hegde, 2017).
5	**Modified IHS Resolution Merge**	This method converts red, green and blue (RGB) bands and transforms into intensity, hue, saturation (IHS) and then the Panchromatic replaces the intensity band (Pushparaj and Hegde, 2017). Following that, an inverse IHS transformation is carried out. The original method has the limitation of fusing only three Multispectral bands but the Modified IHS has the capacity to merge more than three bands (Pushparaj and Hegde, 2017) but the quality is dependent on the spectral overlap between the Panchromatic and Multispectral bands (Zhang and Mishra, 2012).
6	**University of New Brunswick (UNB) sharpening method**	The UNB method is considered to produce a pan sharpened image of very high spatial quality(Sun et al., 2014). In this method, for both the Panchromatic and Multispectral image, the histogram is standardized (Basaeed et al., 2013). The standardization process helps and assures in reducing the spectral distortion when fusing the two images. The spectral overlap of the Multispectral with the Panchromatic image is used to generate the synthesized Panchromatic image. Assignment of band weights is also applied after considering the least square error to produce synthesized Panchromatic image. Finally all the standardized Multispectral band is merged with the synthetic Panchromatic image (Basaeed et al., 2013)

		Image removed due to copyright restriction. (Basaeed et al., 2013)
7	**Ehlers Fusion Method**	The method specifically considers merging of images to preserve the spectral properties of the imageries. The iterative process takes in three bands combinations for colour separation and enhancement using IHS (intensity, hue & saturation) transformation. To decompose the intensity, Fast Fourier Transform (FFT) techniques is used and thus the merged intensity component is generated using information concerning the "low frequency" from the Multispectral imagery and "high frequency" information from Panchromatic imagery. A new IHS image is generated from the merged intensity component and the original image containing saturation and hue components to which then an inverse IHS transformation is performed to the RGB image (Yusuf et al., 2012).
8	**Nearest-neighbour diffusion-based pan-**	The methodology for this method is as shown in the flow chart below which describes the detailed steps that are involved in carrying out Nearest-neighbour diffusion-based pan-sharpening algorithm (Sun et al., 2014). The method is bifurcated into two branches, the left branch basically works using linear regression to generate the digital count contributed by each Multispectral bands to the Panchromatic image. As the Panchromatic and

	sharpening algorithm	Multispectral image have different spatial size and do not overlap it is essential to understand the spectral response function, the Panchromatic image must be down sampled to that of Multispectral image and then perform linear regression. Thus, the vector “T” accounts for the digital count of pixel and corresponding regression error. “T” is also useful in normalising the spectral values. In the right-hand side of the flow chart, the pixel difference factor is introduced to understand the difference of the pixel of interest to that of the nine neighbouring super pixel (in 4*4 grid). If the summation of the difference is zero, it shows the similarity and solution requires high diffusion but on the other hand if the difference is a high value, a more restricted diffusion is suggested. 

Research was carried out on World View-2 imagery and Quick bird Imagery to basically assess the qualities of fused image generated using various Pan sharpening methods namely: Gram–Schmidt, Ehlers, Modified IHS Resolution, High pass filter (HPF) Resolution Merge and Wavelet - Principal Components methods for the assessment based on the various noise-based metrics (Yusuf et al., 2012). In terms of Color reproduction and preservation Pan sharpening methods like Ehlers, Wavelet- Principal Component and Gram-Schmidt methods produced better results. On the basis of noise metrices, Ehlers and wavelet fusion methods yielded good results. On the other hand, when other factors like motion blur and other types of noise were analysed, High pass filter (HPF) Resolution Merge and Gram Schmidt outperformed other methods. The author emphasized the importance of pan sharpening process and the production of fusion image greatly impacted the classification accuracy (Yusuf et al., 2012).

Pan sharpening method has become an inevitable procedure in carrying out image segmentation, image classification and researches that studies the temporal change detection (Jawak and Luis, 2013). Another research was carried out on the WorldView-2 imagery to assess the quality of fused image using various pan sharpening methods. As presented in the above paragraph, the results from this research are also in complete agreement in terms of colour preservation, best results are yielded using Gram Schmidt, Ehlers fusion and wavelet-principal Component sharpening methods. The performance in terms of quality indices were preserved in all the methods used for this study: Gram Schmidt > W-Principal Component > Ehlers Fusion > HPF > Mod-IHS> Brovey Transformation. The author also assessed the accuracy produced using different resampling method with various pan sharpening methods and suggested the use Cubic convolution resampling method in conjunction with Mod-IHS and HPF PAN-sharpening Methods and Nearest neighbor resampling methods to be used in conjunction with Brovey Transformation, Gram Schmidt, Ehlers fusion, and W-principal component PAN-sharpening algorithms. This research further emphasized the statistical inclination and dependency of the pan sharpened image on the choice of resampling method even though resampling method might not depict the visual effect on the fused image but the fused image has to be statistically assessed (Jawak and Luis, 2013).

In article(Sun et al., 2014) the Nearest Neighbour diffuse (NN Diffuse) algorithm was tested on different imageries namely WorldView-2, GeoEye-1 and USGS EO-1 sensors. Small patch of areas like parking lot, residential areas , mall, recycling sites were chosen to implement the NN diffuse algorithm and comparative spatial analysis using other algorithm specifically Bi-cubic ,Gram–Schmidt, genrealized minimum mean square- error (gMMSE) method and University of New Brunswick (UNB) sharpening method. From the reseasrch it suggested that

UNB and NN diffuse method produced high quality fused image with enhanced spatial sharpness. In case of the Gram -Schmidt method, it produced good results but the fused image appeared blurred. The NN diffuse method suffered spatial bleeding in areas where the Panchromatic imagery had lower contrast. The algorithm tested on the GeoEye-1 imagery showed remarkably good result in the complex urban scene and the algorithms showed the correlation of the Panchromatic band with the green, red and near-infrared wavelengths (450-900nm) and least with blue band.

After much assessment and studies on the available pan sharpening methods, researchers have discussed the existence of tradeoffs with respect spectral and spatial components; simultaneously both the spatial and spectral component cannot be incorporate in a fused image without compromising one component, therefore the users has to make a conscious choice in adopting the method that best fits the research area (Marcello et al., 2013).

2.4 Use of Very High-Resolution(VHR) satellite imagery to map informal settlements

In pursuit of mapping buildings using high resolution imagery for the purpose of detecting illegal settlements or for city mapping perspectives, various researches and methods have been undertaken and delved in to resolve the uncertainties and complexities related to such field of interest (Gram-Hansen et al., 2019). Automatic extraction and mapping of building structures from satellite imageries have always been a complex and difficult task which are contributed by ambiguities related to variability in building forms and shapes, obstruction in the imagery caused by trees and skyscrapers, spectral variabilities caused by different roof types because of roof material and its reflectance , spectral characteristics similarities in other land types like road and bare soil with the roof types (Gavankar and Ghosh, 2018, Gram-Hansen et al., 2019).

2.4.1 Object based Image analysis (OBIA) technique

Ever so changing and evolving digital mapping and imaging techniques have been striving for advanced and more intelligent methods for delineation of spatial units with comprehensive integration of knowledge and experiences to understand the complex imaged information (Blaschke et al., 2008, Attarzadeh and Momeni, 2012). Object based Image analysis is one such concept and technique introduced in the field of earth observation which is intended for the automated analysis on the very high resolution imagery and interpret the information from the imaged reality on the basis of spectral , spatial , textural and topological attributes (Blaschke et al., 2008, Nussbaum and Menz, 2008). The guiding principle for OBIA likewise for any other technologies introduced in remote sensing fields is to achieve and extract maximum information from the imaged reality and convey those extracted and best described information to the users (Fallatah, 2020). OBIA utilises the concept of cognitive influence and spatial thinking and considers the basic human perception of environment from the imaged reality. In understanding the complex real world, it requires exhaustive and detailed analysis which are often approached with plurality of solutions and user driven (Blaschke, 2010). The process of OBIA is often considered as cyclic nature that follows an iterative processing rather than a linear one directional input output format. It enables users to accommodate different target classes, flexibility in class modelling, supporting other geospatial data format and data integration and malleable according to the requirements of users (Blaschke, 2010). In OBIA “image object” is the central and basic element for investigation (Nussbaum and Menz, 2008). OBIA encompasses the automated image segmentation which produces segments/ image regions which are plausible with reduced level of complexities and detailing (Norman et al., 2021). Segmentation accounts for grouping of homogeneous neighbouring pixels based on

the shape and spectral properties (Nussbaum and Menz, 2008). Followed by is the image classification which are often referred as knowledge-based classification as those segmented regions are classified into meaningful image objects based on human perception and experts' knowledge for interpretation of the areas(Blaschke, 2010, Nussbaum and Menz, 2008).

Object-based image analysis (OBIA) has become one of the most commonly adopted techniques in building studies which generates segmented features from the amalgamation of pixels into homogenous regions and then classifies the regions into individual image objects (Norman et al., 2021). The accuracy of the classification is strongly dependent on the choice of segmentation parameters and classifiers in generating a qualitative classified image (Norman et al., 2021).

The salient traits a segmented regions shall possess to qualify as an object are (Blaschke et al., 2008):

a. Discreteness: the identified image object possesses the digital number which can facilitate the conversion into vector format for remote sensing related analysis. Proper representation of the smaller objects generalised into one same semantic larger unit but has meaning if it is classified. The trait is linked to the connectedness and unity of the image objects.
b. Coherency: uniformity when classified as an image object but contains variance of spectral and spatial patterns noticeable at different scales.
c. Contrast: to delineate the distinctive entities with differing colour tone, pattern or texture within the regions but appear as same for an image object.

"A good segmentation has little over-segmentation and no under-segmentation" (Belgiu and Drăguţ, 2014) over segmentation occurs when the contrast between the adjacent segments lacks enough knowledge and is rather merged into single objects whereas in under-segmentation, lack of coherency in the adjacent segments leads to the splitting of the segments into separate objects (Blaschke et al., 2008).

Research was carried out for building extraction using Object-Based Classification and complimented with shadow detection analysis for high resolution Multispectral imagery in study area Tetuan, Morocco (Benarchid et al., 2013). The authors used mean shift clustering algorithm for image segmentation and classification was done using Support Vector Machines (SVM) algorithm. SVM Algorithm is a supervised classification technique used for identification of patterns and analyses data based on manually defined training samples for building structures. Building's extractions are definable to certain extent on the basis of the shadow

casted by structures based on the sun's angle/illumination and the shape of the buildings when compared to other land use types. In the Multispectral imagery (RGB Colour space) the shadow information is available in the blue band and the band was explored for the processing of the data and further the intensity-based classification was assisted using Otsu's Thresholding method. The method has the drawback of misclassifying building that doesn't have shadow projection from the classification and thus must be used as restrictive conditions. Various accuracy assessment ratios and statistical methods were used to quantify the buildings captured from automatic approach with the manually digitized building information. The authors discuss the challenges in mapping the buildings in urban areas when compared to rural and sub urban areas and concluded that there were discrepancies in terms of miss factor and overlap factor. The overall detection percentage was 69% and such lower percentages were as a result of spectral similarities in diverse rooftops with the soil and roads and the urban building patterns. The impact of shadow detection method in urban areas were negligible as some buildings and road were misclassified as shadow in some areas.

Another Research was carried out in Jeddah city located in west of Saudi Arabia (Fallatah, 2020) and the author used Object-Oriented Image Analysis (OBIA) approach and OBIA-machine learning to identify and delineate the formal and informal settlements based on specific parameterisation and indices namely :(NDVI - Normalised difference vegetation index, Standard deviation , Visible brightness, NDWI - Normalised difference water index, Border index) and the object segmentation and analysis were carried out in eCognition9 Software. The author stated that the accuracy achieved using only OBIA method resulted in 83% and OBIA using machine learning for larger scale of area resulted in 91%.

A research was carried out in Selangor, Malaysia using Object-based image analysis (OBIA) technique (Norman et al., 2021) and the study aimed at assessing the two classifiers of OBIA technique namely SVM (Support Vector Machines) and DT (Decision Trees) for efficiently mapping buildings using Sentinel-2B imagery in e-Cognition software. The classification was conducted using Machine Learning (ML) algorithms. The authors discussed the importance of setting parameters prior to the segmentation which significantly impacted the accuracy level of the classifications. The overall accuracy achieved using SVM classifiers was 93% in comparison to DT which yielded 73% accuracy.

2.4.2 Pixel based Image analysis

For the pixel-based analysis the essential unit is pixel and the spectral characteristics of pixel is determinant of the classification (Blaschke, 2010). In this approach according to the spectral characteristics of the features the individual pixel in an image is examined, analysed and then classified (Nussbaum and Menz, 2008). Various comparative studies are conducted to assess the level of accuracy achieved in terms of feature extraction using Object-Based Image Analysis and pixel-based image analysis. One such research was out in Qinhuai District of Nanjing, China to extract urban features including road, buildings, vegetation, water and bare land using both methods (Zhang et al., 2019). The overall accuracy achieved using pixel-based image analysis was 64% and kappa coefficient as 0.4889 and on the other hand Object based Image Analysis using visual interpretation yielded an accuracy of 95% and Kappa coefficient of 0.9405 (Zhang et al., 2019). Similar comparative research was carried out using high resolution IKONOS imagery in Safranbolu, Turkey to extract the historical building features (Marangoz et al., 2006). As for pixel-based image analysis Maximum likelihood algorithm was used with some ground truth data and yielded and overall accuracy of 76%. Nevertheless, OBIA resulted in overall accuracy level of 86% and outperformed pixel-based image analysis (Marangoz et al., 2006). The major drawback of analyzing imagery using only pixel-based classification is that in VHR imagery the object to be classified into meaning entity generally is large than the size of pixel which can be a problematic (Blaschke, 2010).

2.4.3 Deep learning algorithms

Automated extraction of buildings using machine learning and other novel algorithms have paved a more promising approach that bridges the gap of applied earth observation and computer vision to deliver the information about the imaged reality (Hoeser and Kuenzer, 2020). The acquisition of imagery and doing manual extraction of building foot print is slow as well as expensive but such challenges can be streamlined by adoption of machine learning approach infused to the earth observation applications (Touzani et al., 2020). Methods adopted for object detection and classification have evolved over the years and with the advancement in use of machine learning that includes increased capabilities in regard to computation and visualization have effectively complimented the object detection (Norman et al., 2021). Convolution neural network which acts as a backbone architecture for performing such object detection have dominated such field of interest (Maxwell et al., 2021a) and detailed introduction is provided in section below.

2.4.3.1 Introduction

Initially as most of the computing was driven by machine learning algorithms which basically includes algorithms like Histogram of Oriented Gradient (HOG) and Shape-Invariant Feature

Transform (SIFT) and produces an input to the classifiers like support vector machines to locate features and classify object on the imagery (Vaidya and Paunwala, 2019). But in the recent years because of advancement in computing technology, a major shift in the adoption of neural network for object detection and classification is prevalent (Hoeser et al., 2020). The state-of-the-art object detection and classification using Convolution Neural network is showcasing remarkable performance (Lee et al., 2016) in terms of handing larger image databases and computation time as operation is facilitated by Graphics Processing Unit (GPU) in powerful computers (Hoeser et al., 2020). The advantages of adoption of deep learning over machine learning are in terms of computing larger datasets, end to end problem solving approach , low interceptibility but often such approach is strongly dependent on the high end machines(computers) and the processing time is longer than machine learning approach (Vaidya and Paunwala, 2019). Deep neural network deals with concealed multiple layers between input and output layers embedding feature attributes that compliments better object detection (Vaidya and Paunwala, 2019). The most popular and widely adopted deep learning technique is convolutional neural network (CNN) (Basri et al., 2018, Hoeser and Kuenzer, 2020).

2.4.3.1.1 CNN architecture

CNN is popularly adopted for object detection as it has capacity to detect edges, corners and much more intrinsic attributes of an object (Basri et al., 2018).

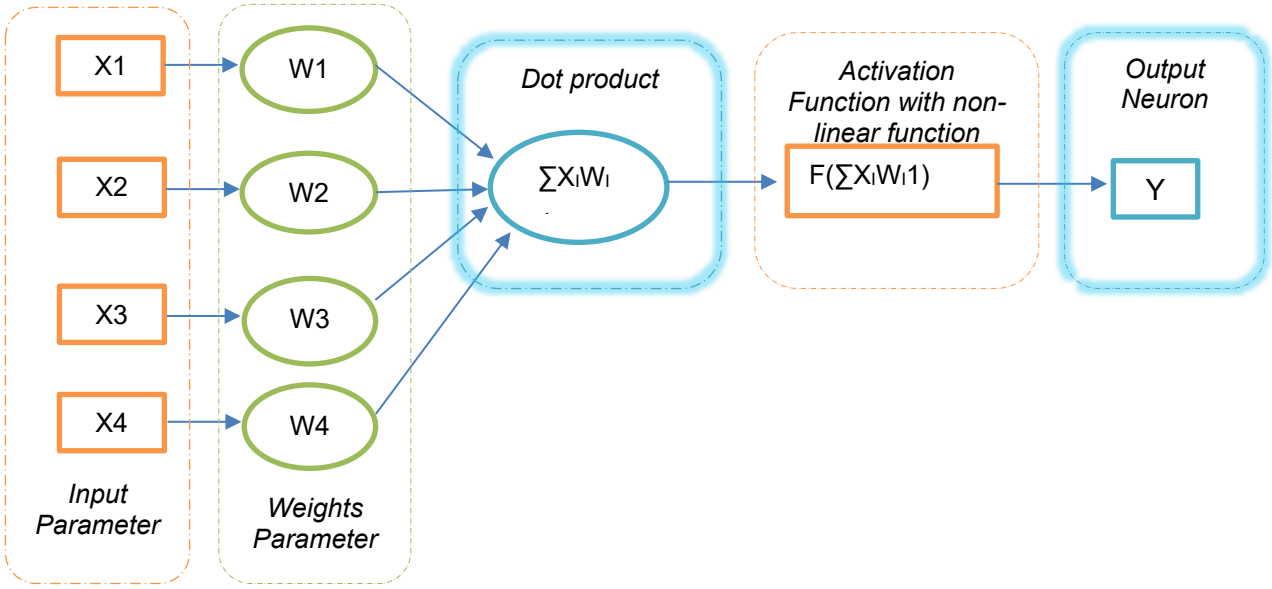


Figure 5. Architecture of artificial neuron (Vaidya and Paunwala, 2019)

Image removed due to copyright restriction.

Figure 6. Detailed CNN Architecture (Vaidya and Paunwala, 2019)

1. **Creating Convolution Layer:** The equation below is used to define the size of the output image and the output image is called as feature detector or feature map. Output image is called feature map. It is an essential step for parameterization to generate an output layer to be used for further processing (Vaidya and Paunwala, 2019).

$$\textit{Equation 1: Size of output image (x, y)} = \left(\frac{n+2p-f}{s} + 1, \frac{m+2p-f}{s} + 1 \right)$$

Where:

a) x, y: size of output image,
b) n, m: size of input image.
c) f: filter size.
d) p: value of padding. This parameter determines the number of pixels that are added to the boundary of an image. Output size decreases without the padding.
e) s: window stride which determines the number of pixels which indicate number of pixels that travels between two windows in both directions.

2. **Pooling Layer:** this process is conducted to reduce the size of the output vector feature for faster computing purposes.
3. **Flattening Layer:** Converts the pooling layer to a single vector format. The output becomes the input vector feature to generate a fully connected neural network.

4. **Fully Connected Layer:** Contains the intrinsic relationship of parameters in the network that can be used for training CNN (Vaidya and Paunwala, 2019).

2.4.3.1.2 CNN Model- ResNet Architecture

The ResNet was introduced in 2015 and have been excelling in serving the purpose for classification and object detection (Hoeser et al., 2020). There are numerous CNN architecture built in namely (LeNet-5, ZFNet, VGGNet, GoogleNet, ResNet, MobileNet) but for the object detection and in terms of percentage of error rate , ResNet outperforms other CNN architecture (Vaidya and Paunwala, 2019). Figure 7 depicts the flow chart of ResNet Architecture.

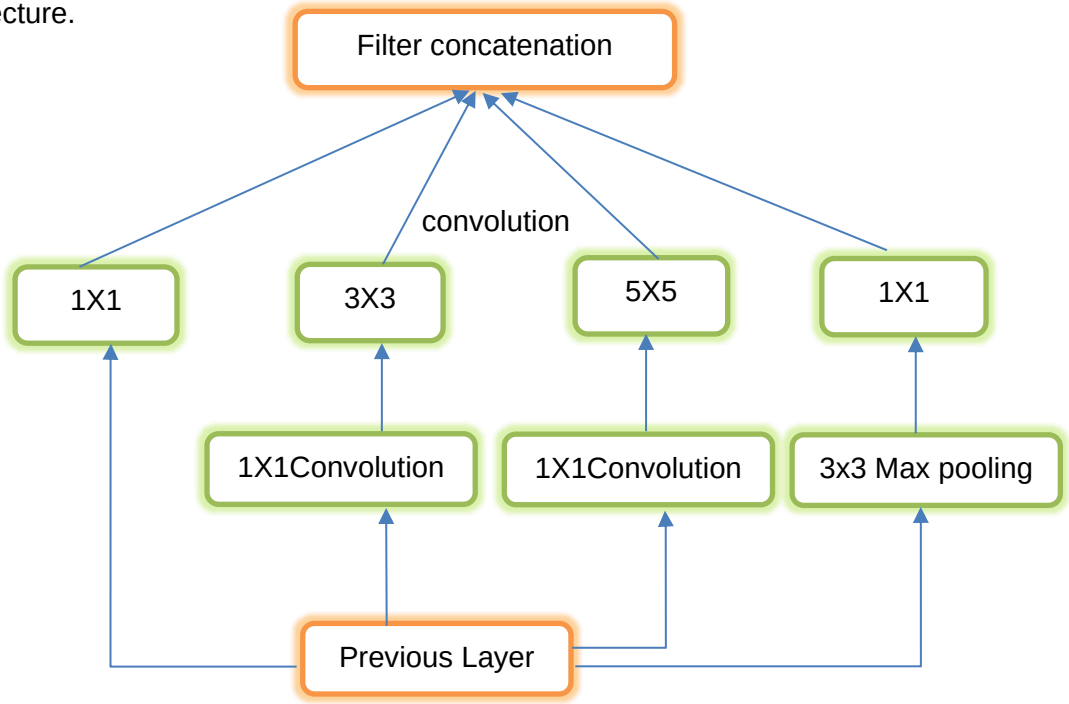


Figure 7. Flowchart depicting the ResNet CNN architecture (Hoeser and Kuenzer, 2020)

2.4.3.1.3 Training of CNN

In this process, the known input values are taken as reference to approximate the location of the target by the selecting the weights from the network generated (Hoeser et al., 2020). It is an iterate and supervised technique wherein the back propagation of error function is applied to the initial weights (Vaidya and Paunwala, 2019). In this part of updating the weights defining the learning rate is essential and needs a careful scrutinisation and iterative processing to deduce an appropriate value (Vaidya and Paunwala, 2019).

2.4.3.1.4 Object detection using CNN

In this section, introduction to various object detection models that are based on CNN architectures are presented:

1. **R-CNN**

It was introduced in 2013 mainly for object detection and classification (Hoeser and Kuenzer, 2020). In this method firstly “region proposal” that is the region of interest containing probable features for extraction are generated. The wrapped region area created considering the initial input image size of CNN from all the sub region proposals (Vaidya and Paunwala, 2019). Thus, the features extracted are utilized for SVM Classifier for classification into the classes defined. R-CNN is time consuming process because of the numerous trainings that are required namely (Vaidya and Paunwala, 2019):

1. Training CNN to extract features from the image
2. SVM Classifier has to be trained that are compatible with the CNN generated features and also on the basis of the classes defined.
3. Region proposal algorithm requires training.

As R-CNN requires training of each region proposal and for SVM classifier , it requires a larger storage capacity of the computer and the training process shall takes days to complete and therefore the detection method is a slower when the features are in validation phase (Hoeser et al., 2020). Nevertheless , this algorithm has the benefit of improving the mean average precision (mAP) approximately by 30% (Vaidya and Paunwala, 2019).

2. **Fast R-CNN**

It is a more time efficient and faster version of R-CNN. Fast R-CNN operates around separate region proposal network (RPN) and produces region of interest. From RPN, it generates the region proposals in the image which are then used as an input for series of convolution and pooling processes (Hoeser and Kuenzer, 2020). The CNN generates a feature map from which feature vector is extracted and saved separately in a special layer called as RoI (Region of Interest) pooling layer (Vaidya and Paunwala, 2019). Furthermore, the output from such conclusion network is taken up for fully convolution network (FCN). From FCN, two major output layers are generated; one is using the SoftMax (is an activation function used for multiple classes classification) that’s helps in predicting the classes, second is regression layer used for prediction of bounding box offset (Vaidya and Paunwala, 2019). Training using Fast R-CNN is much more simpler than the R-CNN algorithms as it doesn’t have to undergo separate SVM classification and the entire training is completed in one go (Hoeser and

Kuenzer, 2020). It reduces processing and training time, resources and storage capacity by nine-fold when compared to R-CNN. The algorithm tested resulted in 66.9% accuracy (Vaidya and Paunwala, 2019).

3. Faster R-CNN

In this algorithm, the base operation and generation of RPN is similar to Fast R-CNN (Vaidya and Paunwala, 2019). But in terms of training network, it deviates from the previous algorithm (Hoeser and Kuenzer, 2020). It is based on switching between the region proposal generation and the object detection and such parallel operation on use of network is realized only on CPU (Vaidya and Paunwala, 2019). Because of that fact, the processing time is much more reduced compared to Fast R-CNN (Lee et al., 2016). It was reported that the algorithm tested on Pascal visual object classes resulted in 66.9% (Vaidya and Paunwala, 2019). Figure 8 depicts the detailed flow chart of faster R-CNN architecture (Veeranampalayam Sivakumar et al., 2020)

Image removed due to copyright restriction.

Figure 8. Flow chart of Faster R-CNN (Veeranampalayam Sivakumar et al., 2020)

4. You Only Look Once (YOLO)

It is CNN based algorithm that doesn't generates a region proposal but rather divides the entire image into numerous sub regions. The class probabilities and bounding box approximation are conducted on every subregion (Vaidya and Paunwala, 2019). For every cell, two anchor boxes are generated with varying size and scale to detect object. The anchor box consist the center of objects and the algorithm calculates two values; one is the class probabilities for classes defined and second is the coordinates of the bounding box (Vaidya and Paunwala, 2019). The accuracy achieved based on pascal VOC 2007 is 63.4% and the accuracy can be enhanced by batch normalization and inclusion of multiple anchor boxes. Smaller objects tends to be neglected in this algorithm (Hoeser and Kuenzer, 2020).

5. Single shot multi-box detector (SSD)

It is a fully convolutional based model like YOLO which helps in detection of object in one pass and doesn't require two stages like R-CNN, Faster R-CNN, Faster-RCNN. The algorithm generates rectangular boxes in the entire image similar to the rectangular grid. Loss function are calculated for the approximated and centered rectangular boxes to the default ground truth bounding box (Hoeser and Kuenzer, 2020). Instead of using the region proposal network it rather divides the region into grid cells that are used for prediction of probable object (Vaidya and Paunwala, 2019). In each grid cell it can contain single or multiple objects with varying scale and aspect ratio, hence in SSD several anchor boxes are assigned for every possible object size during the training phase which matches the anchor boxes with the ground truth object in an image (Vaidya and Paunwala, 2019). The prediction of the object's location and class is based on the maximum overlap of the anchor boxes (Hoeser and Kuenzer, 2020). Values are assigned as 1 for the perfect match and 0 for otherwise (Vaidya and Paunwala, 2019). To avoid the overlapping of bounding boxes non maximum suppression's function are applied (Vaidya and Paunwala, 2019). SSD utilizes a pre-trained backbone model like ResNet to generate a feature map of 256 7*7 given the use of ResNet-34 (Hoeser et al., 2020).

Research has been carried out using Convolutional neural network Deep learning specifically R-CNN method to extract the building polygons. Segmentation was based on pixel wise to mask the building and non-building to generate the initial polygon boundary. The feature extraction using convolutional backbone architecture (ResNet) and classification assisted with the bounding box recognition and mask prediction is applied. The training was phased out in three stages using different epoch and ResNet applied. The 80% of the dataset was trained and 20% used for validation sets. The R-CNN method resulted in 0.717 F1 score which is evaluated based on the algorithm performance as prescribed in Intersection over Union (IoU)

metric (Zhao et al., 2018). As a cost-effective approach, Deep CNN as a part of DeepLabv3 was adopted for mapping informal settlement using Sentinel 2 imagery in areas where the spectral distinction is unclear in both formal and informal settlement. The model used 80% for training and 20% for validation. The model seems to have performed well even in imagery with a lot of noise data (Gram-Hansen et al., 2019) .

2.4.4 Other approaches and algorithms

Differential Morphological profile(DMP) infused with mathematical morphology is another building detector technique which has been widely used for noise filtering , pre-processing , shape and pattern association (Shackelford et al., 2004). The Morphological profile for an image is generated based on reconstruction of opening and closing operators, the building extraction is based on the opening profile and the shadow extraction is based on the closing profile. The algorithm is based on six parameters for analysis and the resultant map is accurate but gives rise to many incomplete building which needs further refinement (Shackelford et al., 2004).

Many algorithms and techniques are being devised and still in the process of developing or improvising to smoothen the process and solve the complexities for such building footprint mapping. Morphological Building index (MBI) is a recent development in automatic detection of buildings using high resolution imagery (Xin and Liangpei, 2012). The original method of MBI leads to omission and commission errors; commission error caused by the similarities in the spectral characteristics of bright bare soil, open areas and roads to that of buildings and omission errors with regard to heterogeneity in the darker roofs, shape or directionality of the buildings when performing the segmentations. But the improvised method in this paper uses the Morphological Building index (MBI) and Morphological Shadow index (MSI) for object-based extraction technique. PanTex is used for processing the imageries as it is considered effective in generating/mapping built up areas for high resolution imageries and SAR images but the optimal image resolution for PanTex is considered as 5m for detecting and mapping built-up areas (Xin and Liangpei, 2012). The joint use of MBI and MSI are constructed on white and black top-hat transformation which are used to identify the bright and dark features in the imagery. Anomalies related to commission and omission errors are addressed in the proposed techniques which incorporates four spatial/ spectral indices for mapping namely MBI (Morphological Building Index), MSI (Morphological Shadow index), GI (Geometrical Index) and NDVI (Normalized difference Vegetation Index). The proposed technique was tested and validated on two high resolution imageries namely IKONOS and World View -2 imageries acquired for Hangzhou, east of China as study area. The research reported that the proposed method successfully reduced the omission and commission errors and the overall accuracy

when compared to the original method; for IKONOS imagery the accuracy was improved from 90% to 94% and for World View-2 imagery the accuracy was improved from 70% to 91%. The proposed method is promising which doesn't incorporate the use of training data for validation and the accuracy generated from applying this method outperforms the methods like Differential morphological profiles (DMP) and recently developed methods called Multiscale urban complexity index (MUCI) (Xin and Liangpei, 2012)

Manual extraction of building footprints is considered to be costly and requires allocation of more resources hence the automated technique have been gaining a lot of adoption and applicability as a cost and time efficient approach (Gavankar and Ghosh, 2018). The authors (Gavankar and Ghosh, 2018) have proposed a method to extract buildings using morphological Top-hat transformation on the IKONOS Panchromatic high resolution imagery to extract the dark and bright features on the imageries. Furthermore, spatial analyst tool like K-means clustering is adopted to determine the centroids for the classes defined (Bright, intermediate and dark building classes) and image refinement is carried out using median filter. The output from the method was applied for both Object based and pixel-based analysis to further validate the performance and accuracy of method. The method proved to have better performance in terms of completeness, quality and correctness for object based than the pixel-based analysis. The overall accuracy for object based technique in terms of correctness is 89% in comparison to 82% using pixel based technique (Gavankar and Ghosh, 2018).

3 CHAPTER: METHODS

This chapter as illustrated in Figure 9, presents a detailed flow chart of the processes conducted and methods involved in identifying the illegal settlements using GeoEye-1 and Pleaides-1A imagery. The methods include the primary data acquisition, image preparation to image analysis and finally accuracy assessment to accurately delineate the illegal settlement.

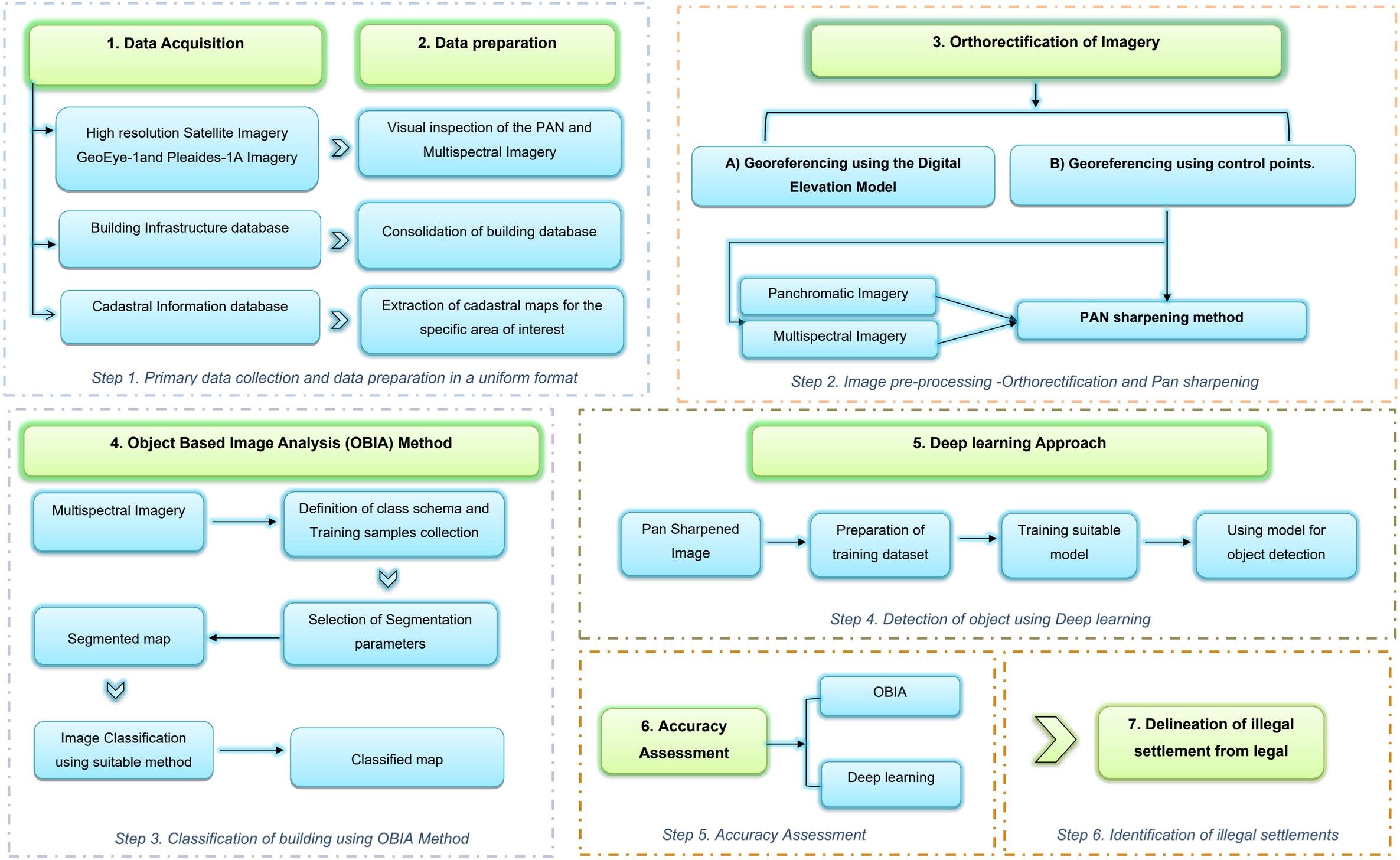


Figure 9. Flow Chart of methods involved in carrying out Object based Image Analysis using Geoeye-1 Satellite Imagery and deep learning approach using Pleaides-1A Imagery to detect objects

3.1 Data Acquisition and data preparation

In section 3.1, presents the various primary and secondary data acquired from relevant authorities and sources to achieve the desired goal of the project.

3.1.1 Satellite Imageries Acquisition

A visual inspection was conducted using google earth interface to capture the dimensions of the minimum dwelling size in the study area to better understand the spatial requirement of the satellite imageries to map the smallest building units for the study area. Google earth is one of the virtual globes available that provides free accessibility to larger sets of remotely sensed images to facilitate a base map or assist in ground validation in many instances (Yu and Gong, 2012). The positional accuracy of the images is varying and mostly high RMSE are observed over developing countries. But nevertheless, the images available in google earth are sufficient and plausible enough for validation and reference purposes (Yu and Gong, 2012)

Figure 10. Visual assessment in defining the minimum dwelling unit size using google earth

A very high-resolution satellite imagery requirement was inferred after a visual inspection to map the minimum dwelling unit as depicted in Figure 10.

There are numerous optical imageries available and in Figure 11, it depicts the broad categorisation of imageries and its optical specifications are illustrated. The use of the imageries is highly dependent on the applicability and field of interest.

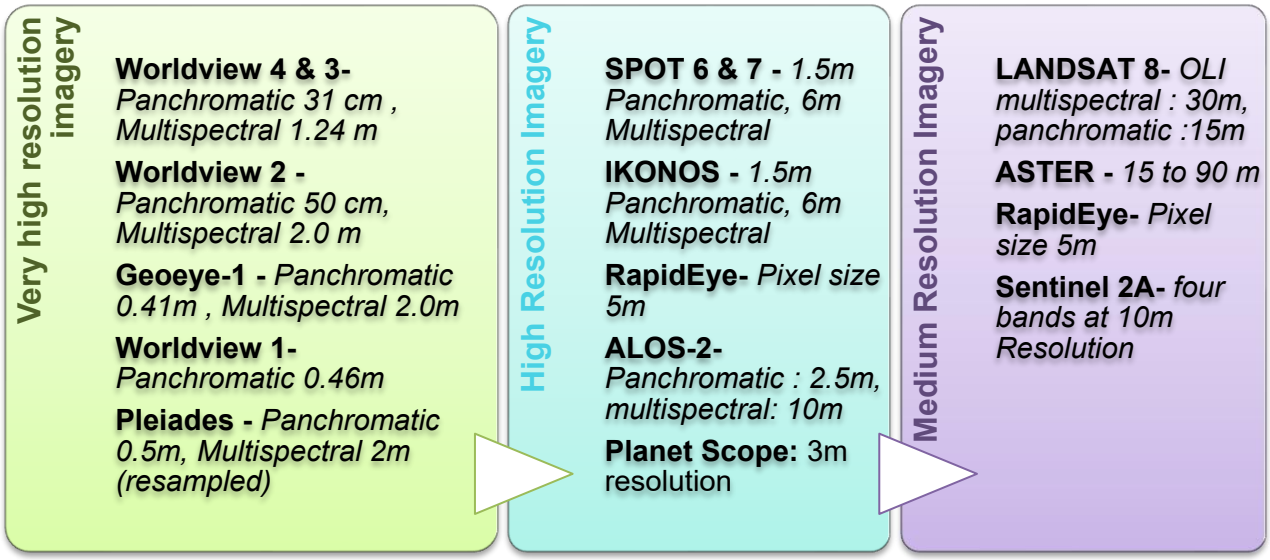


Figure 11. Optical imagery specifications (Acker et al., 2014)

3.1.1.1 GeoEye-1 Imagery

GeoEye-1 Imagery was launched in September 6, 2008 at Vanderberg Air Force Base, California delivers Panchromatic imagery of spatial resolution of 0.46 meters and Multispectral imagery of 1.84m (Jacobsen, 2011). GeoEye-1 has four band (Panchromatic, red, green and blue band). The satellite revisits lesser than 3 days and have the capacity to locate objects within three-meter distance as well. The satellite is flown at an elevation of 681km and can detect object with diameter equal to or more than nadir ground sample distance of 46 centimetre. The Nominal swath width of GeoEye-1 imagery scene is 15.2 km at nadir (Jacobsen, 2011).

Upon consultation with the relevant authority (National Land Commission) in Bhutan, GeoEye-1 imagery was acquired for the area of interest. The following Table 3 presents the acquired image specifications:

Table 3. Acquired data- GeoEye-1 Imagery details for the study area

Specification	Details
Creation Date	03/19/13
Acquisition Date/Time	2013-03-18 04:52 GMT
Customer Project Name	Samtse
Scene area acquired	185 sq.km
Sensor Type	Satellite
Sensor Name	GeoEye-1

Datum	WGS84
Pan Cross Scan	0.43 meters
Pan Along Scan	0.43 meters
MS Cross Scan	1.73 meters
MS Along Scan	1.73 meters
Scan Azimuth	272 degrees
Percent Cloud Cover	4

3.1.1.2 Pleaides-1A Imagery

Pléiades1A was launched on December 17, 2011 in Kourou, French Guiana and delivers Panchromatic Imagery of resolution of 0.5m and Multispectral imagery (RGB) at resolution 2m. The Pléiades1A has four bands (Panchromatic, red, green and blue band). The satellite revisits on daily basis and acquires approximately an average of 500,000 sq.km per day. The nominal swath width of imagery is 20km and the product Pleiades comes with automatic orthorectification (Space, 2021).

For the research the Pléiades1A imagery for the scene (80sq.km) as illustrated in Figure 13 was acquired for time period 5/11/2017 on 09/10/2021. The data was purchased from this website (https://www.intelligence-airbusds.com/).

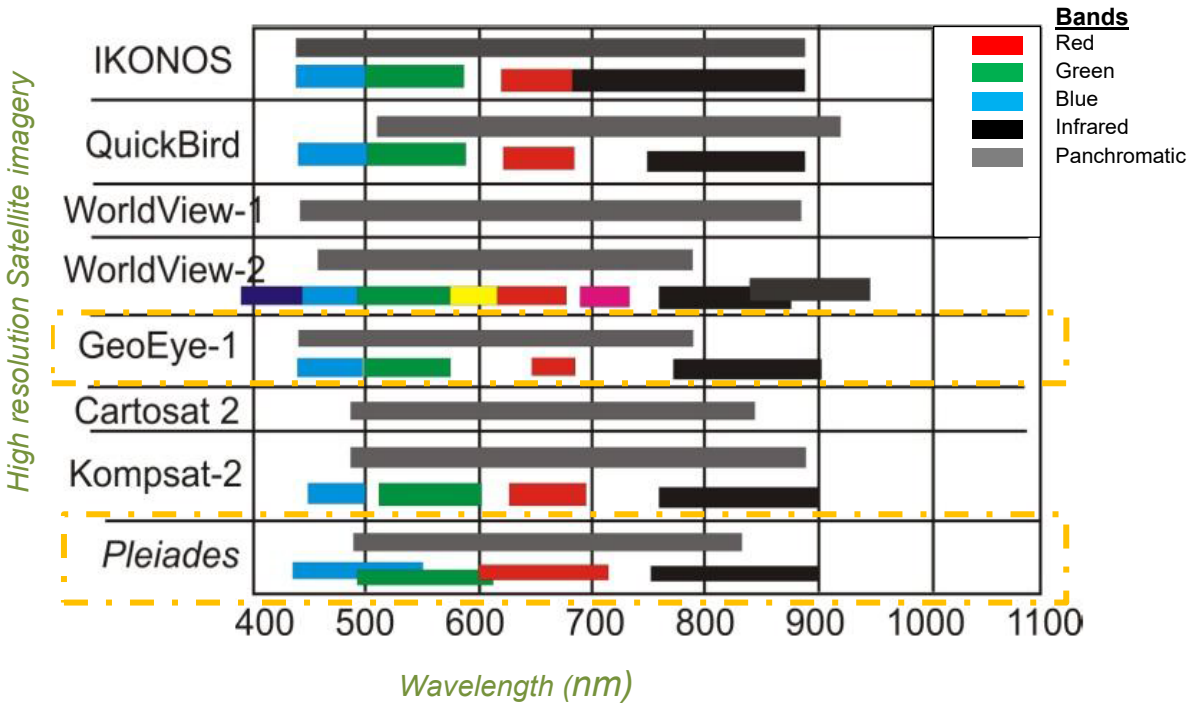


Figure 12. Band width Specification for GeoEye-1 and Pleaides -1A Imagery (Space, 2021)

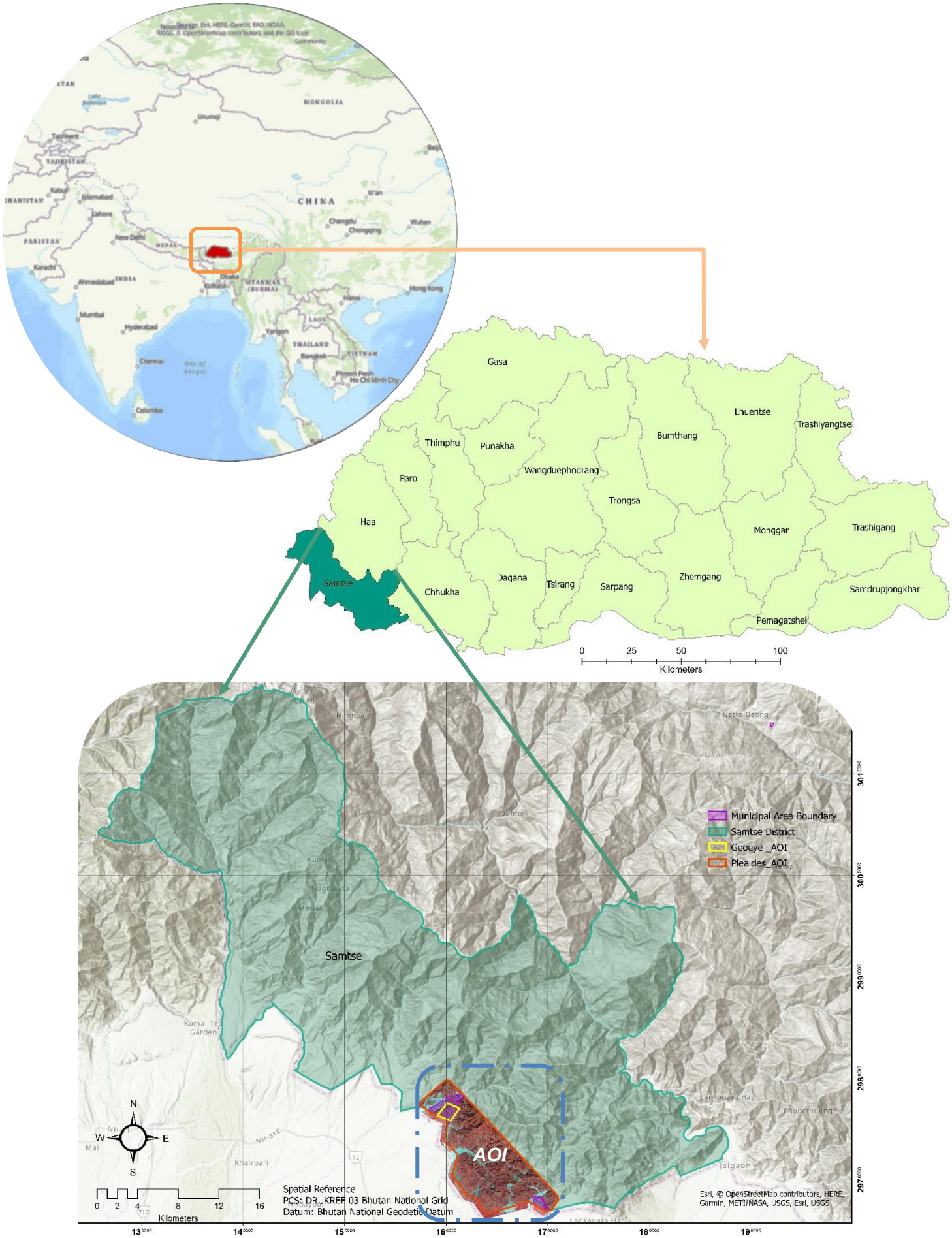


Figure 13. Location map for acquired imageries for area of interest

3.1.1.3 Building Infrastructure database

For the Southern belt of Bhutan under the JICA (Japan International Cooperation Agency) project, topographical base mapping was carried during the year 2014 till 2017 and the project also collected data concerning the infrastructure locations which also captures the various building point and polygon information. The data was thus acquired from the government which will be an important procedure for ground truthing and accuracy assessment.

3.1.1.4 Cadastral database

Tracing back to the olden days, the documentation of land holdings in Bhutan was solely carried out for the taxation purposes and a very rudimentary approach of using standard gauge or Gunter's chain was used in those days to measure the size of the land holdings. By in by with the advancement in the surveying field, even in Bhutan it witnessed a paradigm shift from using manual based estimation approach to more advanced and high precision surveying approach (NLC, 2012).Bhutan transitioned from using plane table to now adoption of high precision and positioning Global Navigation Satellite System (GNSS) for surveying cadastral plots and mapping purposes. Over a very short span of time land prices increased exponentially and prominent upheaval of land tenureship and conflicts became a very concerning and sensitive issues and it is even to this date. Under the command of his Majesty the King of Bhutan, a resolute endeavor was undertaken by the Land department of Bhutan to resurvey the entire land parcels in Bhutan using GNSS approach in the year 2008. The primary goal of the resurvey programme was to strengthen land tenureship and to resolve land issues and conflicts. The cadastral database is constantly updated with the land parcel consolidation and fragmentation cases (NLC, 2012). In the Figure 14, depicts the different land uses under the privately registered land parcels where approximately 3% of the of the land holdings are under residential use and rest are either for cultivation or for other land uses (NLC, 2012). Land for human settlement is very scarce and constitution of Bhutan mandates 70% of the land area be to conserved as forest and for ecologically sensitive areas (MoWHS, 2016).

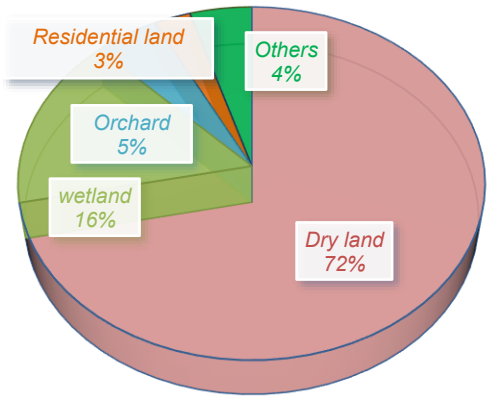


Figure 14. Land Typologies for registered legal plots (NLC, 2012)

3.1.2 Data Preparation

It is an essential step in preparing data in prescribed format before any treatment is applied on the original imagery and database available. For this project, the data preparation was carried out for the building infrastructure database because the data provided had building information both in point and polygon format. For the uniformity in the data format, all the building polygon data were generated as point data by calculating the centroids of the polygon. Then all the point data were consolidated in one layer and extracted for the area of interest. All the data type to be used for the study area are projected in Bhutan projected coordinate system:

- Projected coordinate system: DRUKREF 03 / Bhutan National Grid
- Projection: Transverse Mercator
- Datum: National Geodetic Datum
- Ellipsoid: GRS 1980
- Unit: meters
- False Easting: 250000.0
- False Northing: 0.0
- Central Meridian: 90.0
- Scale factor: 1

3.1.2.1 Software used

For the Data preparation, image processing and geospatial analysis the software used was Earth Resources Data Analysis System (ERDAS) Imagine 2018, and ArcGIS Pro 2.8 were used.

3.2 Visual Assessment and interpretation

Upon the visual investigations, few findings were made as follows that needed to be addressed in adopting better methods and incorporation of pre-processing treatment before carrying out analysis.

a) Co-registration error

When overlaying the Multispectral imagery and Panchromatic imagery, the imageries had significant co-registration error as the images weren't aligned properly. The product came in with standard georectifications but cause of such error is still ambiguous but further processing had to be done before considering for further analysis.

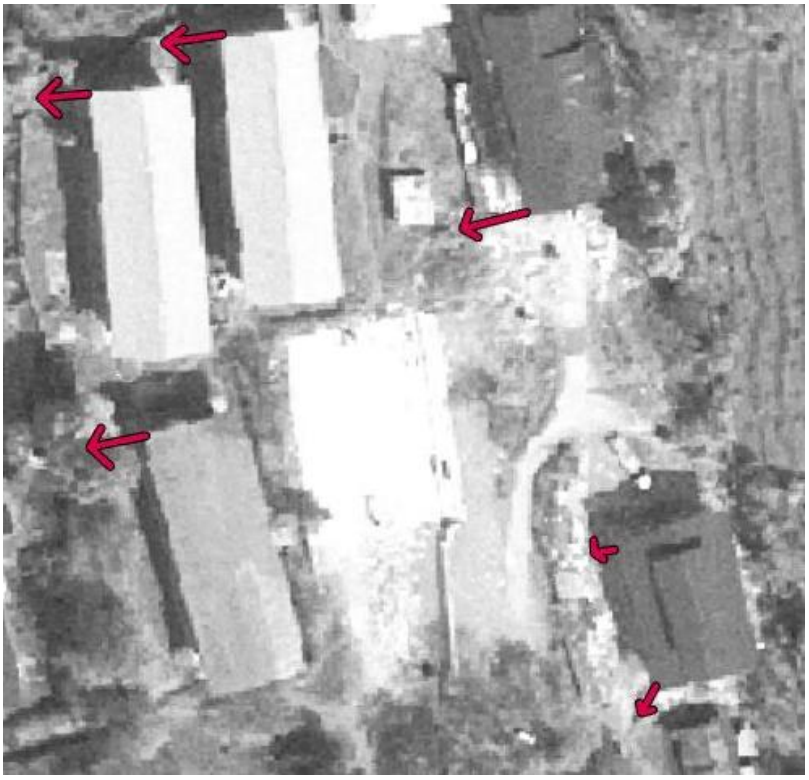


Figure 15. The arrow representing the alignment error of Panchromatic imagery with the base imagery

b) Building database not exhaustive

The building database acquired after scrutinisation with the GeoEye -1 Imagery, it was observed that there was absence of numerous building points which are most likely a building feature. So, a most appropriate method must be adopted to build a comprehensive building database for the purpose of accuracy assessment.

c) Spectral overlap and similarities

It is common and basic understanding in remote sensing that all surfaces have differing reflectance and emission behaviour to the electromagnetic wavelength. Such distinguishing behaviour by different surfaces are reported as spectral signature and the signature are significantly influenced by the material of the reflecting surfaces, topography and atmospheric modifications (Nussbaum and Menz, 2008). For a better understanding, the spectral signature of the distinct surface was manually plotted against the reflectance value using the spectral profile tool in ArcGIS Pro.

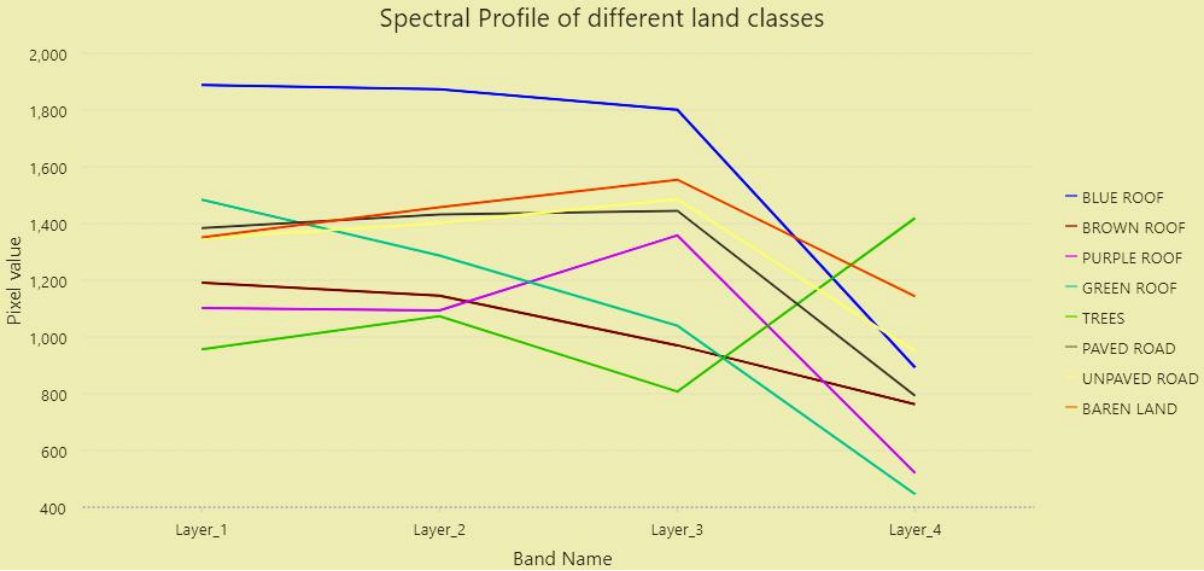


Figure 16. Spectral profile of land classes

3.3 Image preprocessing - Orthorectification of Raster images

3.3.1 Overview on the process of orthorectification

Orthorectification is a process that converts the images into a mapping format that removes the distortion likely to be caused by the imaging sensor, distortion due to earth curvatures, and topographical relief and approximation of geoid which are corrected in the ground control segment but at times flexible for further rectification by the users as well (Chmiel et al., 2004). Georeferencing is one such method of orthorectification that's encompasses the identification and alignment of geographic objects or data to a known coordinate system by the ways of topographical features of areas, by aligning the geometrical layouts of the objects or the semantic properties of object mainly for visualisation and analysis purposes (Hackeloeer et al., 2014). For any applications such a land use land cover mapping , image segmentation and classification, georeferencing an image in the GIS platform, Orthorectification is an inevitable and a very essential procedure as a part of image pre-processing (Aguilar et al., 2017). It enables users to perform orthorectification using raw VHR imagery with ancillary data such as Digital elevation models (DSM) and Ground control points (GCP) (Toutin, 2004, Aguilar et al., 2017). Ground Control points are the survey points of known location points (Aguilar et al., 2012) or can be the conspicuous and identifiable landmarks which are selected from the imagery to correlate with the base imagery and in some cases like for flight based imaging system, it is referred to as the pre-defined or post marked on field for geo-referencing purposes (Hackeloeer et al., 2014). Some products are available with already orthorectified and has high accuracy but that has higher price attached with such products (Chmiel et al., 2004). The newer additions from the Digital Globe Satellites (GeoEye-1 Imagery) and

Worldview Imagery presents Panchromatic imagery with GSD (Ground sampling distance) lower than 0.5m and the dataset comes with accurate geolocation of 5m variability without Ground control points (Aguilar et al., 2017). There are various approaches available for Orthorectification as follows (Aguilar et al., 2017, Aguilar et al., 2012):

1) Use of Ground control points (GCP's) obtained from using Differential Global Positioning System (Real time kinematic mode) on the ground and rectifying the images based on the control points for checking the positional accuracy.

2) Using Google Earth Digital Globe which extracts GCP'S from google earth with direct Geolocation approach. The strategy is useful in Orthorectification for those sensors which consists of relatively better geolocation without ancillary GCP'S like in case of GeoEye-1 or Worldview-(1,2,3) sensors. As those sensors have spatial resolution lower than 0.5m , the residual error can be rectified with simple 'X", and "Y" shift (1st order polynomial) (Aguilar et al., 2013).

3) Using Google Earth which extracts GCP'S from google earth without direct Geolocation approach. It is an approach recommended for VHR imagery with no direct geolocation and containing high order errors in the imagery requiring complex shift and correction model (Aguilar et al., 2017).

An improvised orthorectification can be conducted by including the Digital elevation model for better accuracy and smoothness. Digital elevation surface models can assist in rectifying the Panchromatic imagery. But it was an important observation stated by the authors that the impact of using DSM to generate orthoimageries might be small and lead to issue of non-smoothness while using Nadiral VHR imageries (Aguilar et al., 2017). Given the use of global DEM, it might generate a smooth area but it shall have variabilities considering the morphology of mountainous and complex urban areas. The accuracy level depends on the attributes of site areas, imagery used and the method to orthorectify the imagery (Aguilar et al., 2017). The reliability of the orthorectification can be further enhanced by foregoing the Two dimensional rectification which might not rectify the distortion related to relief displacement and thus demanding 3D geometric correction (Chmiel et al., 2004) ; using the Digital Elevation model for Z coordinate (Aguilar et al., 2013). The method of using Google Earth for positional accuracy was considered to have variabilities and even reported in numerous researches that Ground control points from Google earth approach in developing countries showed poor accuracy (Aguilar et al., 2017). Google Earth engine can serve as a virtual platform in providing georeferenced images and as a source of extracting Ground control points but the

positional data is based on approximation and lack of documentation has led to many researches in determining the accuracy of Google Earth imageries (Aguilar et al., 2017).

Accuracy assessment approach for validating the output from orthorectification was mainly by conducting an independent check using check points not inclusive in the Ground control points and generating the RMSE or using illustration of vector format discrepancies for detailed understanding (Chmiel et al., 2004).

3.3.2 Study area specific orthorectification Methods-GeoEye-1 Imagery

As illustrated in Figure 17(A) using ArcGIS Pro, the raw Panchromatic imagery was projected to the desired location which was more of an approximation method to align the imagery in reference to the base imagery. This was an essential step before carrying out the detailed orthorectification procedures.

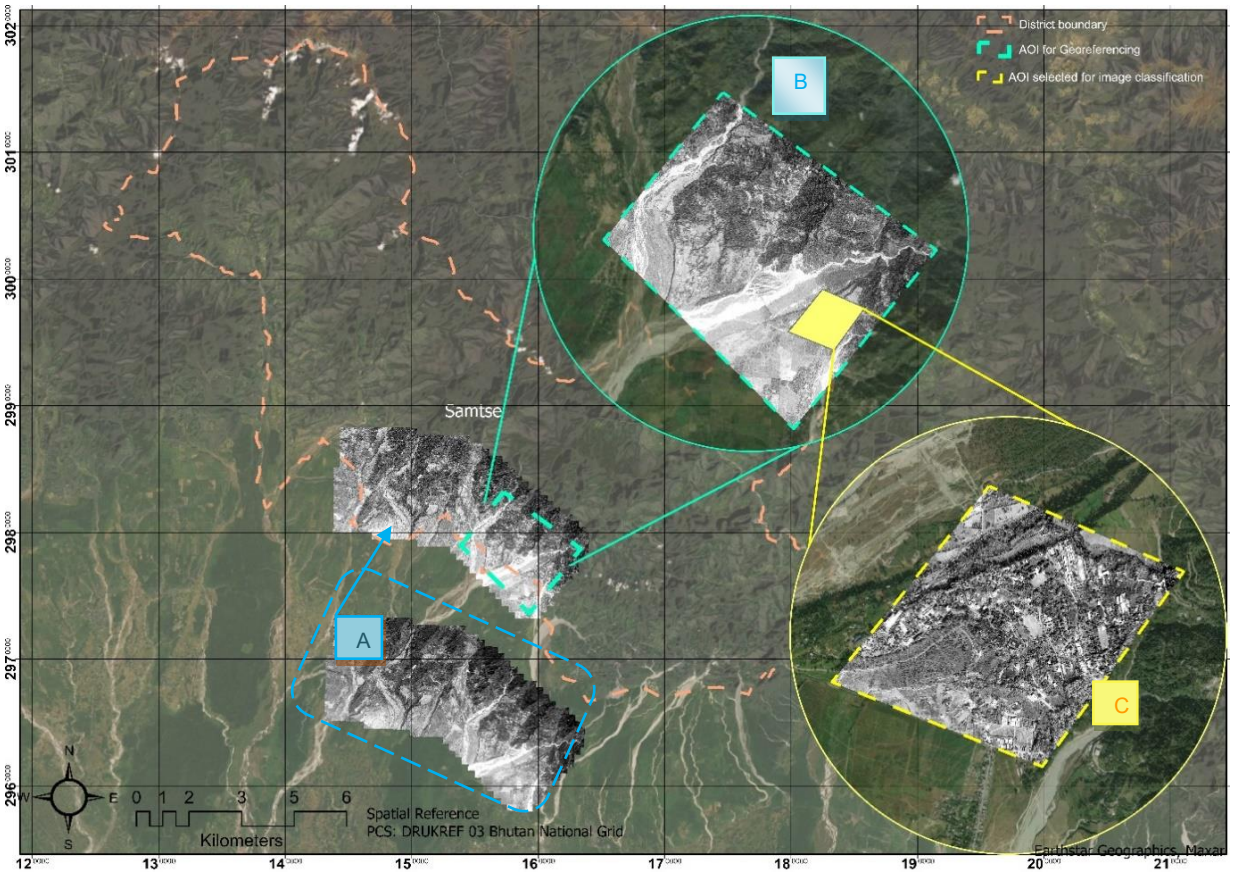
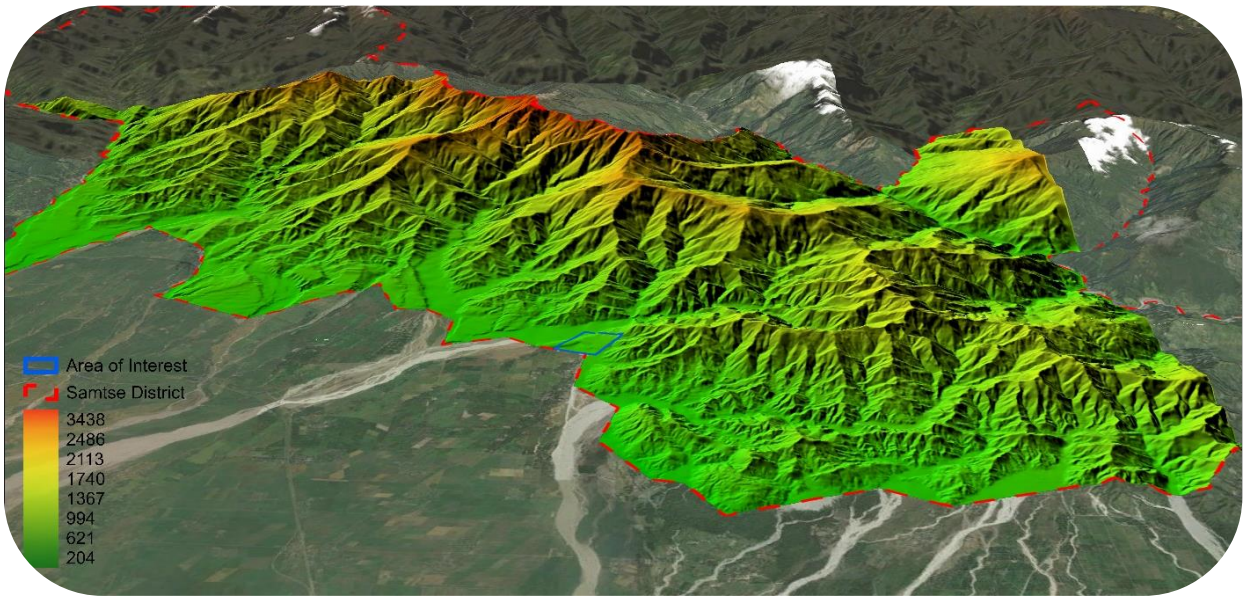


Figure 17. Defining area of interest considering the extent of processing and anomalies related to georeferencing

For the study area, two methods were adopted for the orthorectification as follows:

3.3.2.1 Orthorectification using the Digital Elevation model (DEM):

Topographical base map data was acquired from the concerned authorities and the survey was conducted in 2016-2017 for the southern belt of Bhutan for the scale of 1:50,000. The contour interval for the given data was approximately around 30m and thereby optimal resolution of 10m Digital Elevation Model was thus produced as illustrated in Figure 18. The processing was carried out in ArcGIS Pro software and for the generation of model; TOPO to raster tool was used. Using the Ortho Mapping tool in ArcGIS Pro, the orthorectification was carried out taking DEM model generated as one of the input data. As depicted in Figure 17 (A), the Area of interest was used for the geo-referencing purpose. The southern part of Bhutan falls within the southern foothill region and highest elevation in the district is approximately around 3439m and lowest elevation is around 205m.



. Figure 18. Digital Elevation Model Generated for the study region on the basis of topographical base map for the Southern belt

3.3.2.2 Extraction of control points using Google Earth engine

A. Georeferencing the Panchromatic Imagery
B. Georeferencing the Multispectral Imagery taking Panchromatic imagery as Reference map.

After careful visual scrutinization, it was observed that the imagery had significant co-registration error for both the imageries and much complex shift and ample control points was required to rectify such discrepancies. Therefore, a new area of interest was defined to carry out the process of georeferencing in more detailed manner as illustrated in Figure 17(B).

The Georeferencing process was carried out in ArcGIS Pro Software, by manually defining the control points from the Panchromatic imagery and referencing it to the base imagery. The software enables matching (Hackeloeer et al., 2014) where the user can manually orient the location of desired object from the Panchromatic imagery to the supposedly target location in the base imagery. For the project, 625 Control points were extracted by considering the identifiable and distinct feature locations like road intersection, building corners, road bends and Mouth of roads. The extraction of control points are for building polynomial transformation which shifts and correlates the raster layer from the current location to the spatially accurate target location (ESRI). To ensure better alignment and area of overlap for georeferencing , the extraction of control points were widespread across the study area rather than confinement in one area(ESRI) as depicted in the Figure 20 below.

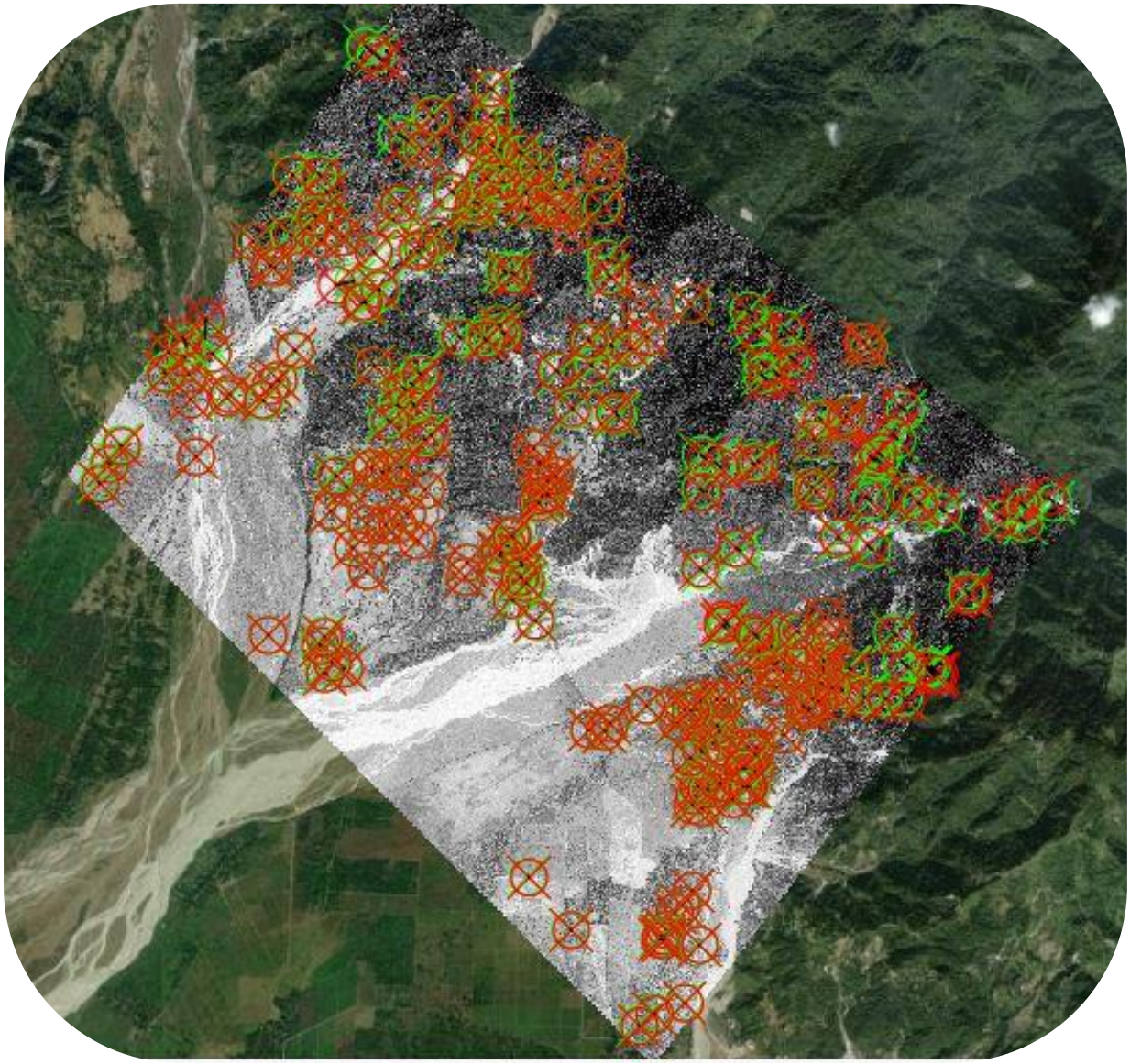


Figure 19. Illustration of creation of control points on the Panchromatic imagery to the target location of base imagery for georeferencing

As illustrated in Figure 17(C), a smaller patch of area (1.6*1.4 km) was thus defined to further curtail the co-registration error and test the method of image segmentation and object classification.

Using ERDAS software the geo-referencing of the Multispectral imagery was carried out taking Panchromatic imagery as the reference map. In the ERDAS software, the polynomial geometric model was selected for geometric corrections. The output file was defined in the projected coordinated system.

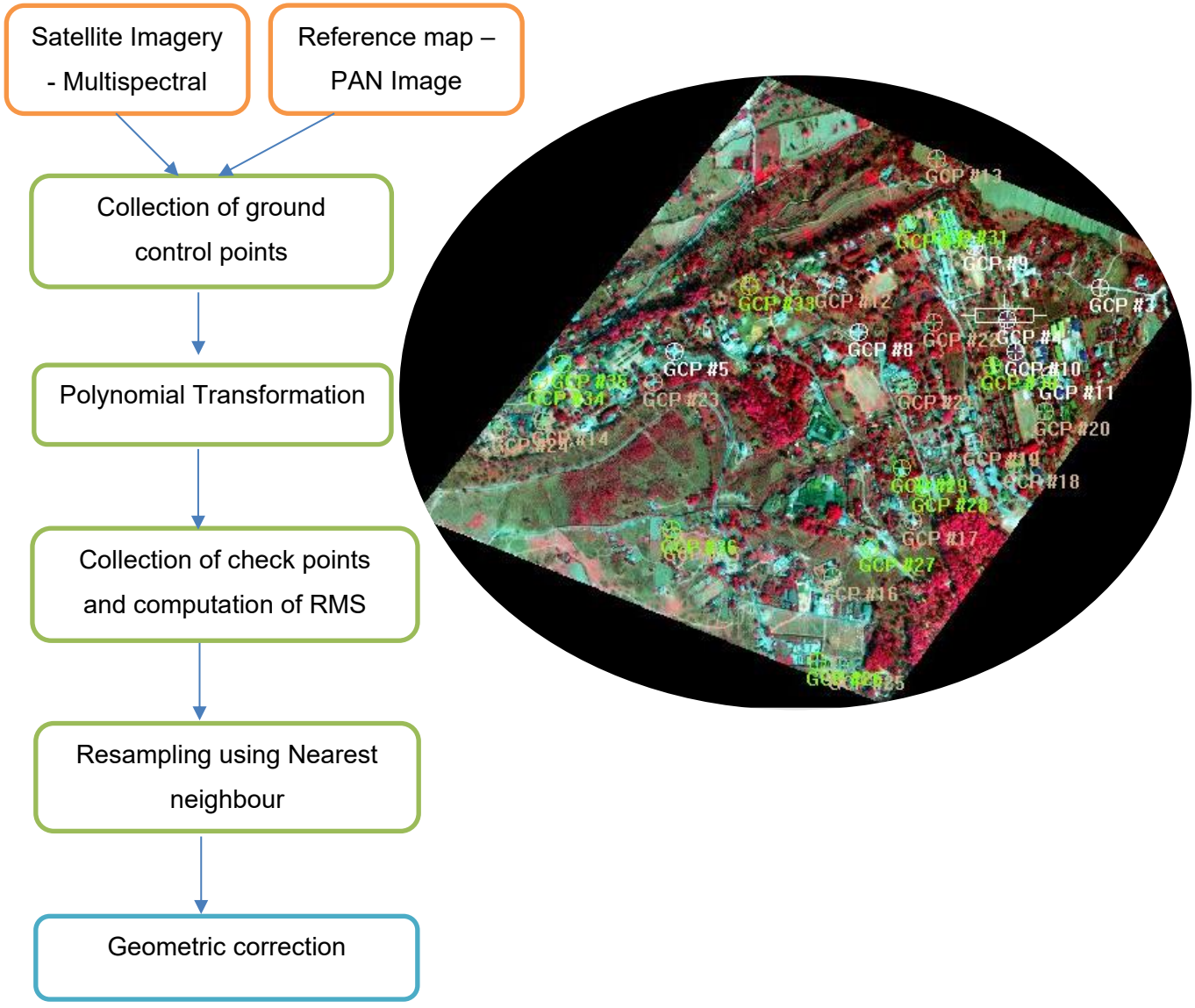


Figure 20. left) Flow chart illustrating orthorectification of Multispectral imagery in ERDAS software, right) Extraction of control points from Multispectral imagery for Orthorectification

3.3.3 Study area specific orthorectification Methods-Pleaides-1A Imagery

A simple 1st order polynomial transformation was applied and approximately 646 control points were extracted on the Panchromatic imagery in ArcGIS Pro for orthorectification as illustrated in Figure 21. Then using ERDAS Imagine geometric correction on Multispectral imagery was carried out in reference to the orthorectified Panchromatic imagery as illustrated in Figure 22.

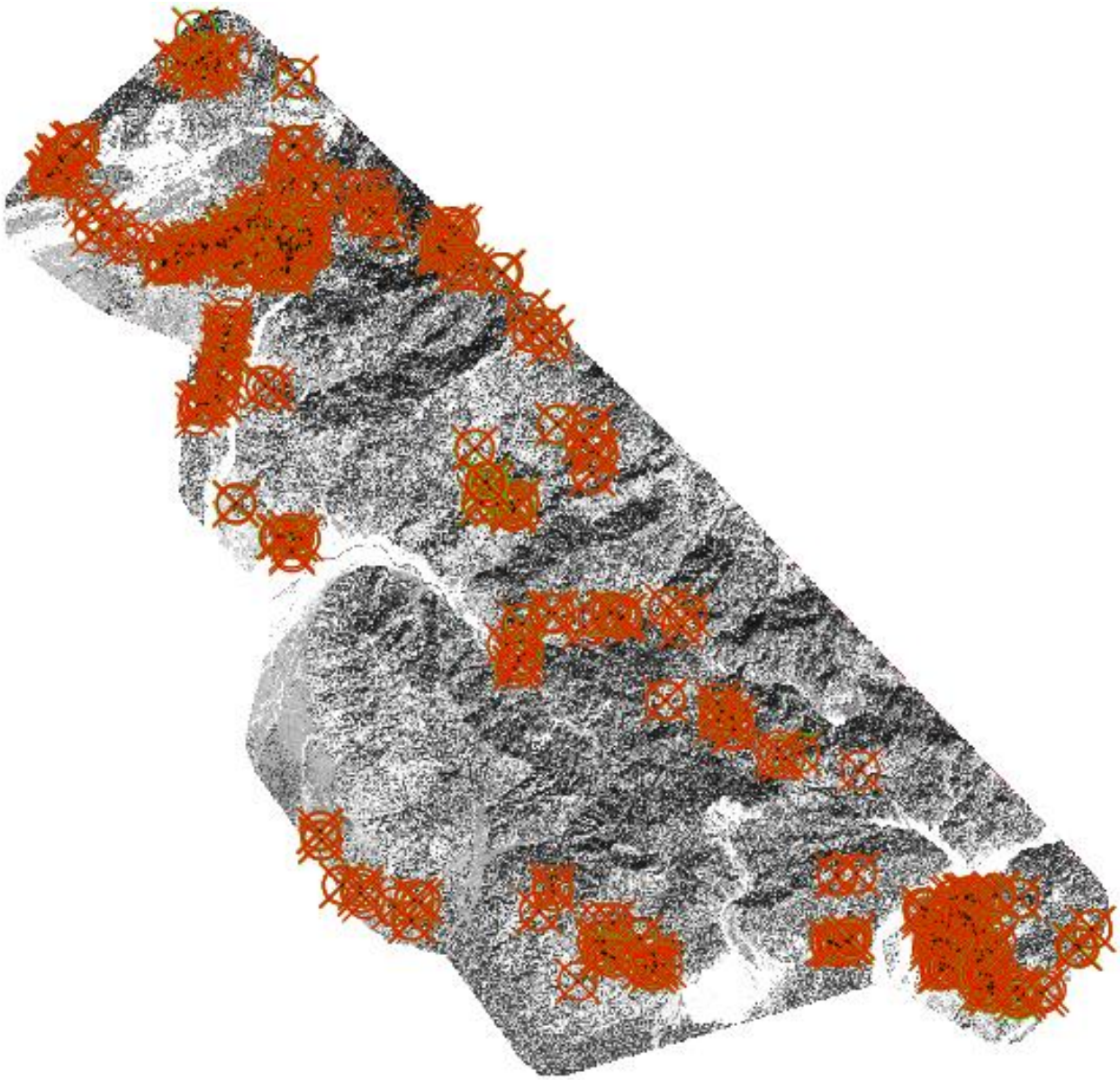


Figure 21. Orthorectification using control points on Panchromatic imagery in ArcGIS Pro

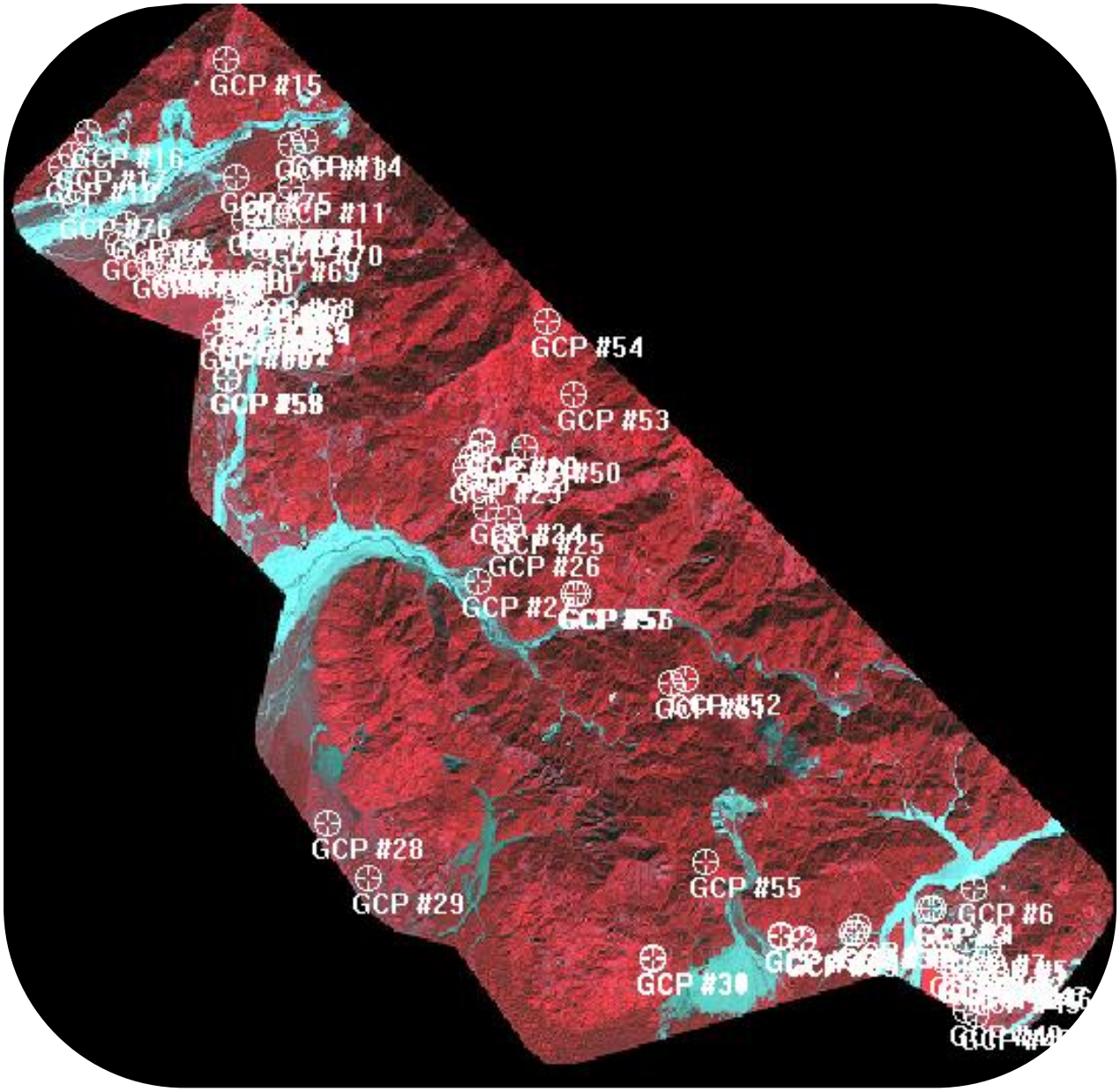


Figure 22.Geometric correction of Multispectral imagery in ERDAS Imagine

3.4 Pan Sharpening of raster images

The pan sharpening of the imagery was conducted for both the imageries in ERDAS software using Nearest Neighbour Diffusion method. The choice of such method was after an intensive literature review on the various methods available as presented in section 2.3.

3.4.1 Pan Sharpening for GeoEye-1 Imagery

The Pan sharpening method for GeoEye-1 imagery was not conducted because of significant co-registration error but OBIA was conducted on Multispectral imagery.

3.4.2 Pan Sharpening for Pleaides-1A Imagery

Figure 23, illustrates three Pan sharpening methods investigated namely: Gram Schmidt, Nearest Diffusion method and IHS resolution merge method. Using ERDAS Imagine software, Nearest neighbour Diffusion method was selected for the study area that suit the complex Urbanscapes and scope of research.

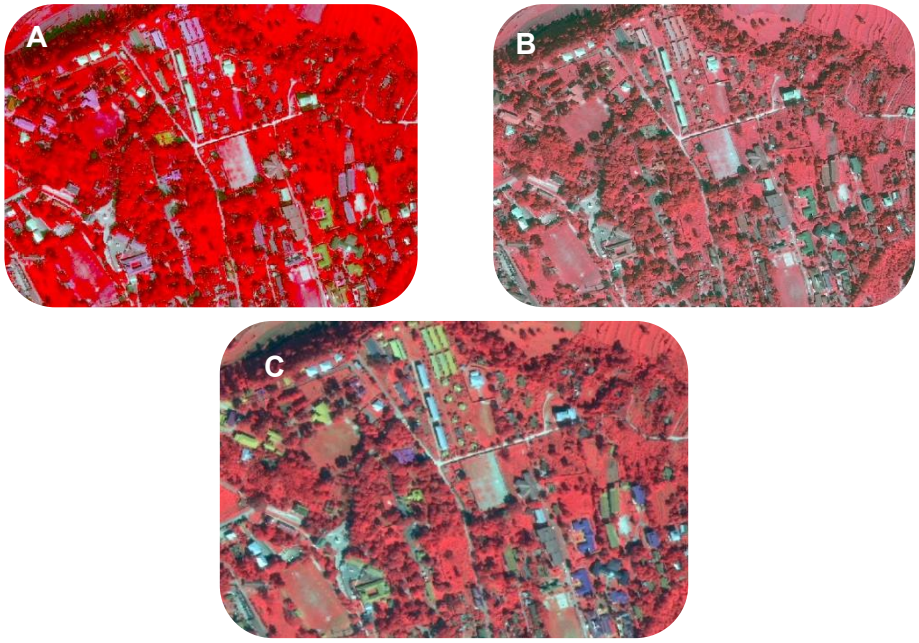


Figure 23.Image illustrating the three-pan sharpening method applied on the imagery and depicting RGB colour band with 2 standard deviation contrast stretch A) Pan sharpened image using Gram Schmidt method B) Using IHS Resolution merge, C) Using NN Diffuse method

3.5 Object classifcation and detection model

In this section it presents the methods involved in image classification using OBIA method and object detection using Deep learning specifically Single Shot Multi-Box Detector model. Object based Image Analysis (OBIA) Method-GeoEye-1 Imagery

Figure 24, depicts the flow chart of processes involved in carrying out image segmentation and classification using OBIA method using GeoEye-1 Imagery.

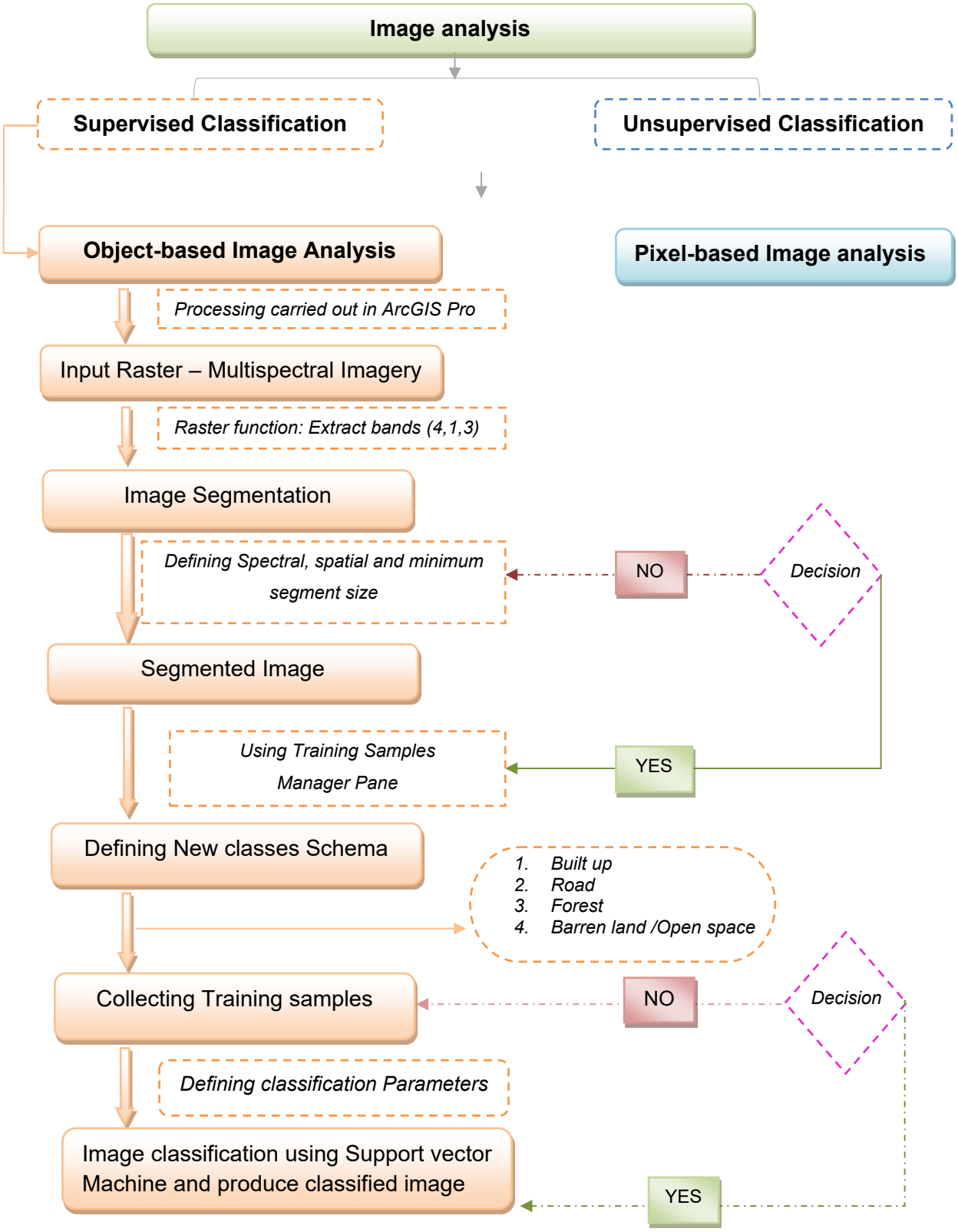


Figure 24.Flow chart involving methods for image segmentation and classification using OBIA method

As illustrated in Figure 24, the image classification can be broadly classified into supervised and unsupervised classification. Supervised classification is a user defined approach where the user defines the land class categories and parameters for segmentations and enables user to redefine the misclassified objects to the target classes (Nussbaum and Menz, 2008). On the other hand , Unsupervised classification is a computer based approach where the classes are assigned based on the spectral properties of the pixels/region (ESRI).

For the study area Supervised image classification Using Object-Based Image Analysis was adopted. The processing was carried out in ArcGIS Pro software and following are the detailed explanation to the methods is illustrated in Figure 24.

3.5.1.1 Supervised Image segmentation

Segmentation is the primary stage for carrying out object based approach wherein the pixels in the image are grouped to form an meaningful individual objects (Attarzadeh and Momeni, 2012) and thus segmentation is based on the desired parameters defined by the users (Nussbaum and Menz, 2008). The level of accuracy achieved for classification is highly dependent on the quality of segmentation and comparatively supervised segmentation gives better result for image classification (Belgiu and Drăguţ, 2014). The segmentation follows three basic criteria namely the homogeneity within the segmented image, the separability from the neighbouring segments and homogeneity in shape of the segments (Nussbaum and Menz, 2008). Various softwares provides different segmentation algorithms but as the ArcGIS Pro software was used for research, following parameters were available with the inherent principles of segmentation maintained. The following parameters needs to be set to find a sound ground which distinguishes the object better for detection.

a) **Spectral details:** value ranges from 1-20 which determines the importance given to the spectral differences of the object in the imagery.
b) **Spatial details:** Value ranges from 1-20 wherein the importance is given to the object proximity.
c) **Minimum Segment size:** Minimum Mapping unit expressed in pixel.

For the study area after numerous segmentations attempts and visual assessment, following values were defined for further processing and the segmented map is as appended in Annexure A.

a) Spectral Details: 8
b) Spatial details: 17
c) Minimum Segment size: 20

3.5.1.2 Training samples for classification

After generating a segmented image, using Training Sample Manager pane available in ArcGIS Pro, it enables the users to define a new classification schema based on the land classes available in the respective imagery. After defining classification scheme, numerous training samples were collected based on the segmented image.

Figure 25.Training samples collection for different land classes from the segmented image

The land classes were broadly categorised into four types and the primary focus was the collection of building data for the classification.

Table 4. Land classes definition and Training samples details

SL. No		Defined Land classes	Specification	Total Number of training samples
1		Built up	Building structures irrespective of precinct use and the materials of the structures. Objects with regular shape and form that can be distinctively identified from the segmented image.	410
2		Barren land / Open space	Barren land, paddy fields that are distinctive from the forest cover, Open space that includes playground and parks	28
3		Forest cover	Trees and densely vegetated areas	51
4		Road	Both paved and unpaved road, includes footpath that are distinct in the image.	21

3.5.1.3 Supervised image classification

The following classifiers listed are widely used Classifiers for object detection and ArcGIS software facilitates the use of Random Forest (RF) and Support Vector Machine (SVM):

a) Random Forest (RF): Using Bootstrapped sampling method, multiple decision trees are created (Kavzoglu et al., 2020). To create the decision tree model, it randomly selects two third of the training data from the total training samples and remaining samples are used for validation of the tree model (Kavzoglu et al., 2020).
b) Support Vector Machines (SVM): It is a non- parametric classifier which performs the classification on those areas where the pattern between different variables is unknown. It is also applicable on those areas which show nonlinear and data with multiple classes(Kavzoglu et al., 2020). Several researches have reported that classification for building extraction (Norman et al., 2021, Belgiu and Drăguţ, 2014) and land use classification (Kavzoglu et al., 2020) using SVM classifier have yielded high accuracy than other classifier.

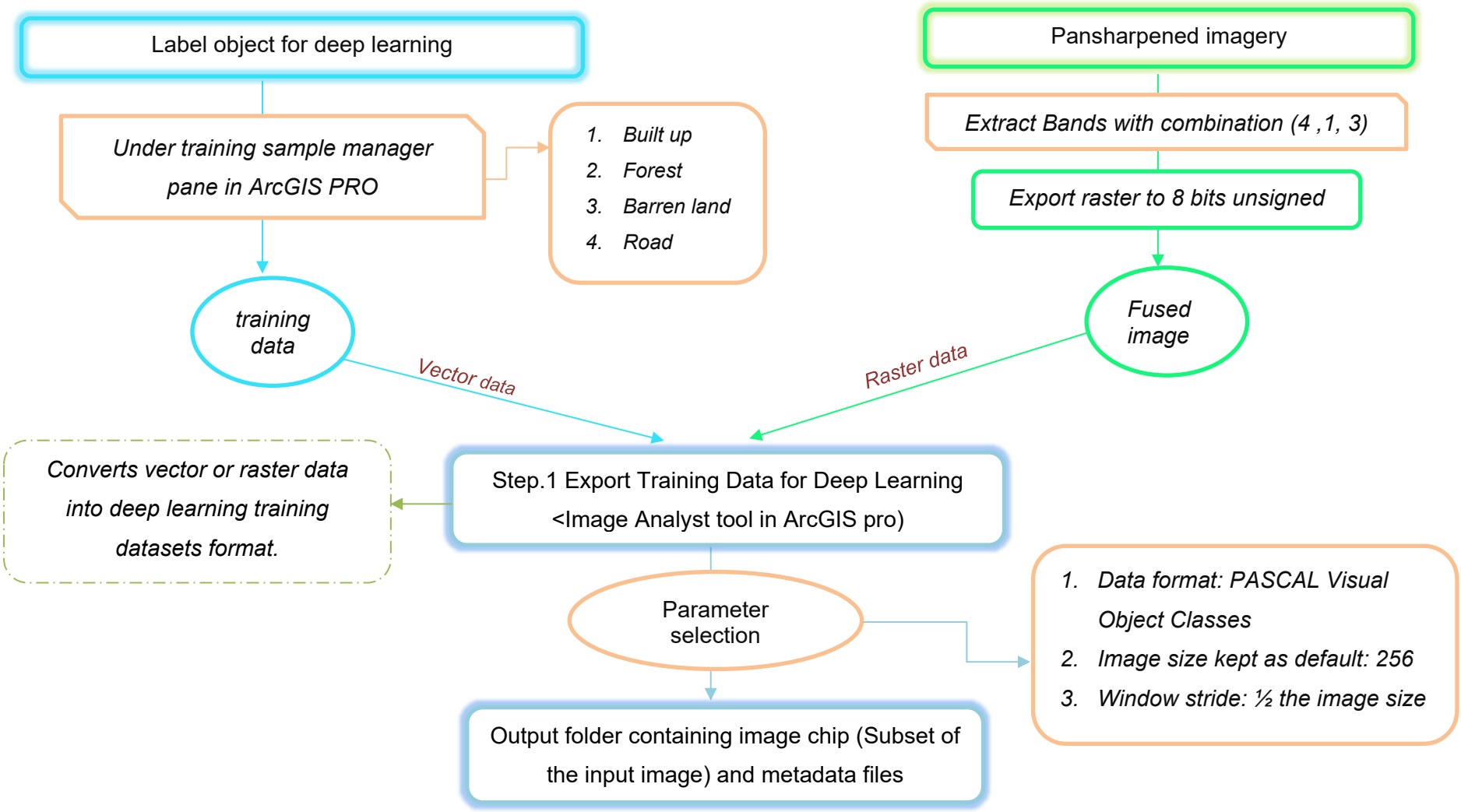
For this research, SVM classifier was used to carry out Object based Image Classification. The classifier considers the compactness and the pixel count, rectangularity of the features to define classify the objects.

3.5.2 Deep learning approach- Pleaides-1A Imagery

For this research, the deep learning model selected was Single Shot Multi-Box Detector model and Figure 27 depicts the overall methods involved for carrying out the object detection model.

3.5.2.1 Deep learning using Single Shot Multi-Box Detector (SSD) for the study area

Following method were involved in carrying out deep learning using SSD deep learning model and the processing was carried out in ArcGIS pro using the deep learning packages available



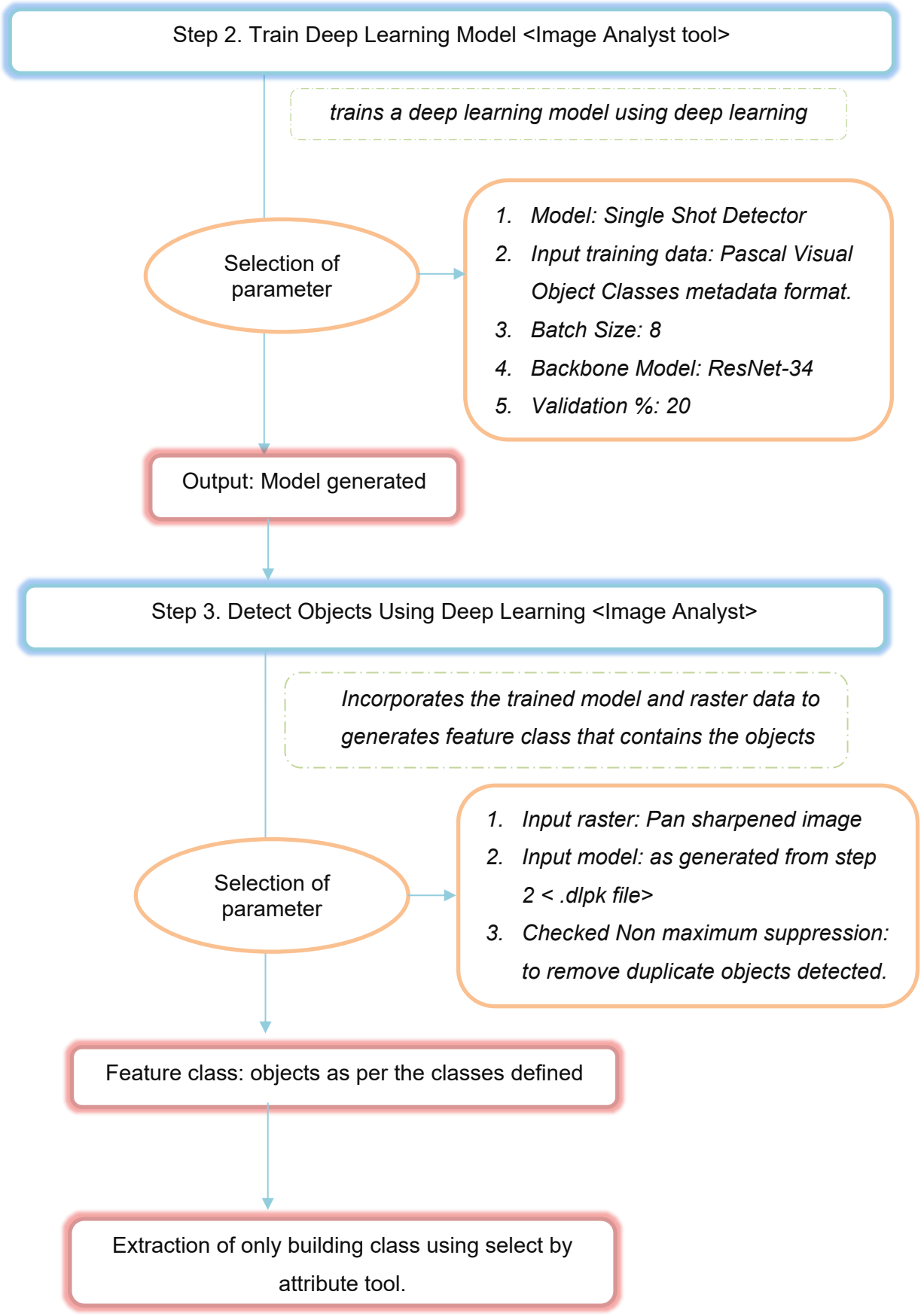


Figure 26. Workflow of deep learning approach using SSD Model

3.6 Accuracy assessment

Accuracy assessment is an essential process to assess the quality and accuracy of the image classification/ detection model and output generated from various algorithms to detect the desired features/objects (Belgiu and Drăguţ, 2014). The assessment considers the reference dataset which can be in various format (ESRI). This section would detail out the various accuracy assessment techniques adopted to assess the classification and object detection output.

3.6.1 Accuracy assessment for classified map using OBIA method

In ArcGIS Pro Software, under Image classification Wizard, after the collection of training samples for defined land classes the user is automatically directed to the selection of random points pane. The optimum number generated by the system is 500 and the user can choose to define more based on the sampling techniques to be used and number of classes defined for the project. There are three types of sampling technique available in ArcGIS Pro Software, namely: a) Stratified random sampling which creates points and the distribution is random relative to the study area, b) Equal Stratified random sampling: creates point with equal representation of each classes and c) Random sampling where points are created and distributed randomly (ESRI).

3.6.1.1 Overall Accuracy assessment using Confusion Matrix and Kappa coefficient

Various studies pursuing Object based Image analysis to detect building footprint using high resolution Satellite imagery for determining the overall accuracy of the classified image, the confusion matrix and kappa coefficient is adopted (Belgiu and Drăguţ, 2014, Benarchid et al., 2013, Norman et al., 2021, Fallatah, 2020).

3.6.1.2 Quality assessment and Accuracy check metrics

Many studies to assess the quality of the building classification, have adopted the approach to create the reference dataset by manually digitizing the buildings identifiable from the imagery and comparing with the detected buildings based on various statistical parameters/ metrics which have been referred in numerous literatures (Benarchid et al., 2013, Gavankar and Ghosh, 2018, Shackelford et al., 2004). Before applying the quality metrics, the basic reporting on the three categories of classified building data generated from the model have to be completed to further assess the quality of classification (Gavankar and Ghosh, 2018, Benarchid et al., 2013). The categories are as illustrated below in Figure 27, the Venn diagram shows the building detected from semiautomatic or automatic approach, building digitized manually using other reference imageries and the building which are undetected using both the approach in the area.

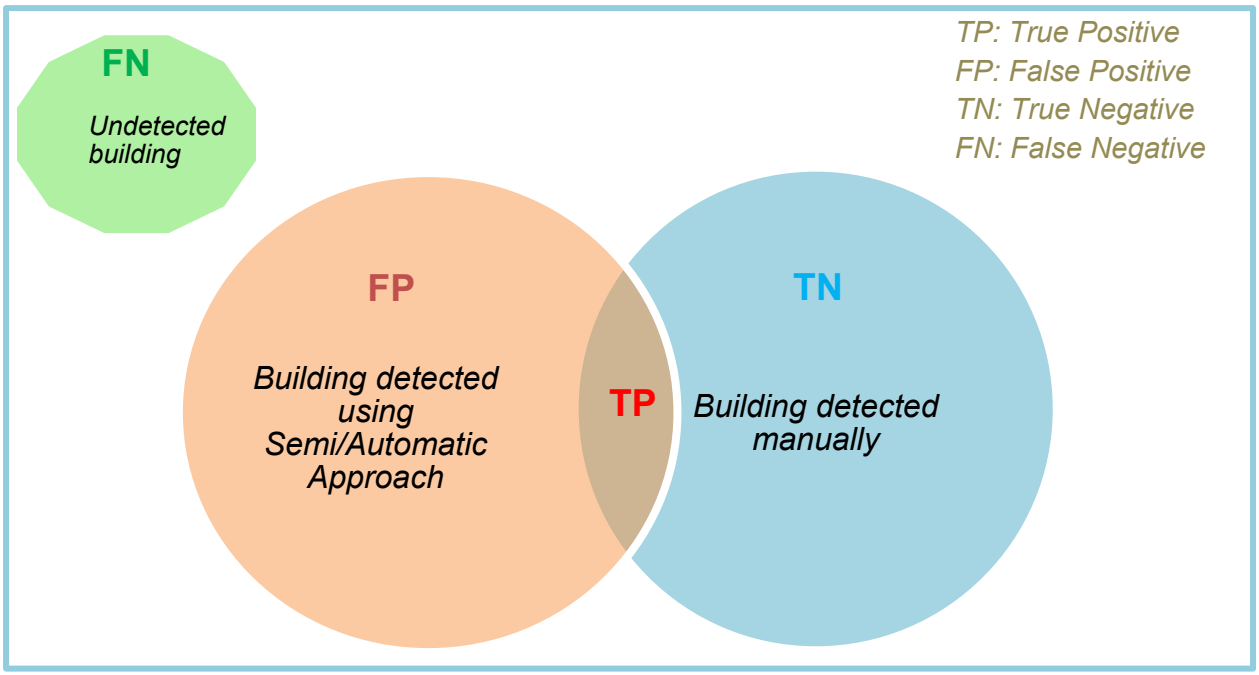


Figure 27. Venn Diagram depicting the three categories of building detected from OBIA approach and extracted from digitization.

I. True Positive (TP): Total number of buildings detected using both the approach of manual digitization and semi/ automatic approach.
II. False Positive (FP): Building detected by only the semi/automatic approach (Benarchid et al., 2013). Falsely extracted as building but not building and is an omission error (Gavankar and Ghosh, 2018)
III. True Negative (TN): Building detected by only manual digitization.
IV. False Negative (FN): Commission error: the total number building undetected by both approach but is a true building but not extracted (Gavankar and Ghosh, 2018).

3.6.1.2.1 Area level assessment using Quality metrics

After the categorization of the detected buildings, the quality metrics are devised and adopted in various researches to assess the quality of output and classification model performance (Benarchid et al., 2013).

I. **Branching Factor**: this ratio depicts the degree of over detection of non-built-up areas as building.

Equation 2: Branching Factor = False Positive/ True Positive

II. **Miss factor:** ratio depicts the omission error in detecting the buildings.

Equation 3: Miss Factor = False Negative / True Positive

III. **Detection percentage:** percentage of building detected using Semi/automatic approach.

*Equation 4: Detection Percentage = True Positive / (True Positive +False Negative) *100*

IV. The **Quality Percentage:** ratio depicts the quality of the detection model.

*Equation 5: Quality Percentage = True Positive/ (True Positive + False Negative + False Positive) *100*

3.6.1.2.2 Object level Quality assessment metrics

The rationale behind carrying out the object level assessment using various metrics is to further authenticate the misclassification of buildings(Shackelford et al., 2004, Benarchid et al., 2013). The objective is to investigate and identify the true buildings left unextracted (FN) , non-built up area misclassified (FP) and validate the true building Detected(TP) (Gavankar and Ghosh, 2018). The acceptable percentage overlap of detected buildings (TP) in reference to the digitized building map is 50-70% (Gavankar and Ghosh, 2018)

The two metrics are considered as object level assessment metrics (Gavankar and Ghosh, 2018, Benarchid et al., 2013, Shackelford et al., 2004):

1. **Correctness:** Detected buildings (DB) which are partially overlapping with the digitized building data taken as Reference Buildings (RB) dataset expressed in percentage.

Equation 6: Correctness = Detected buildings/ Reference buildings

2. **Completeness**: Percentage of Reference buildings (RB) which are in compliance with the detected buildings (DB).

Equation 7 : Completeness = Reference Buildings/Detected Buildings Validation of the Building database for the study area

The acquired building database from the relevant government contains essential information associated with building use and location but the data requires amalgamation in a common format as buildings data are represented as point features in some areas and as polygon features in few areas. This process is necessary to authenticate the viability of using such database which shall ultimately assist in accuracy assessment. Centroids for the polygon feature is created to consolidate the building information in point feature format.

As a part of cross validation process, the manual digitization of building features is conducted by referencing to Multispectral imagery of GeoEye-1 Imagery.

3.6.2 Accuracy assessment for Objects detected using Deep learning approach

Unlike the OBIA approach, reporting the accuracy on the basis of confusion matrix and overall accuracy is not plausible (Hnatushenko et al., 2018). However, it can be reported based on the accuracy indicators/metrices like True Positive (TP), True Negative (TN), False Negative (FN), and False Positive (FP) and other metrices like Precision, recall and F1 Score for more clarification. For conduction of such metrices it requires the ground truth as one dataset for validation and for this research, manual digitization of buildings was conducted by populating the digitized polygon upon visual assessment using pan sharpened imagery.

In ArcGIS Pro, Compute Accuracy for Object Detection tool enables to compute accuracy and the generate report to determines the performance of model. The computation is based on the Intersection or Union (IoU) Threshold and based on that other accuracy metrices are generated (ESRI, 2021, Padilla et al., 2020). The concept of IoU implies the overlap of the detected objects with the ground truth object to determine the performance of the model. The value ranges from 0-1, where higher the value would indicate higher overlap and therefore higher precision of the model (Padilla et al., 2020)

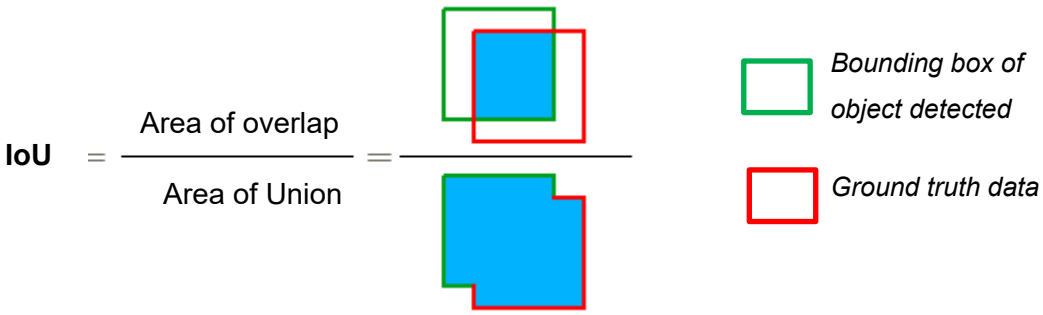


Figure 28. Concept of Intersection or Union (IoU) algorithm (Padilla et al., 2020).

3.6.2.1 Model performance indicators/metrices

For assessing the performance and accuracy of the single class object detection model following indicators are generally used (Park and Kim, 2020, Padilla et al., 2020):

1. True Positive (TP): Correct detection and agreement of object detected and ground truth with specified parameters.
2. False Positive: Incorrect detection caused by the misclassification of other objects/classes.
3. False negative: Objects which remain undetected.

4. True Negative: Objects remain undetected but is a true object. TN is not considered for the model as it tends to create many bounding boxes that has no similarity with the ground truth data.

IoU assists in classifying the so called “correct” and “incorrect” detection that measures the overlapping area(User defined threshold) of the ground truth and bounding box of the object detected to (Padilla et al., 2020).

5. **Precision:** Ratio of true positives to the total prediction

Equation 8: Precision = (True Positive)/ (True Positive + False Positive)

$$\frac{TP}{All\ detection}$$

6. ***Recall:*** *Ratio of true positives to the total number of ground truth objects.*

Equation 9: Recall = (True Positive)/ (True Positive + False Negative)

$$\frac{TP}{All\ Ground\ Truths}$$

7. **F1 Score:** Weighted average of precision and recall and the value ranges from 0-1 which implies 1 as the highest accuracy.

Equation 10 : F1 score = (Precision × Recall)/ [(Precision + Recall)/2]

8. **Precision recall curve:** determines and evaluates the performance of the object detection model. The precision is plotted in y-axis and recall in x-axis. An object detection model is considered good if precision remains high with increasing recall which implies less false positives detected (Padilla et al., 2020). In the Figure 29 illustrated below it shows an instance of the precision-recall curve (Maxwell et al., 2021b)

Image removed due copyright restriction.

Figure 29. Example of a Precision -recall curve (Park and Kim, 2020)

Interpretation of the Precision-recall curve are described below with two indicators (Padilla et al., 2020):

1. Average Precision (AP): average of precision value ranging from 0-1 as interpreted under the precision-recall curve. There are two methods of calculating AP, one is using 11 points interpolation and then drawing the PR curve (Padilla et al., 2020, Park and Kim, 2020), second is calculation for all point which are available in ArcGIS Pro deep learning packages (ESRI, 2021). As the curve is normally zig zag, interpolation helps in approximating the area for better interpretation.
2. Mean Average Precision(mAP): Average of AP interpreted under multiple IoU threshold defined. AP considers a curve generated for one class but for multiple classes and for multiple thresholds designed, mAP is applicable (Park and Kim, 2020, Maxwell et al., 2021b).

In the output from object detection model, the objects have associated confidence level which represents: 1) probability for the existence of the object and 2) delineation of object and degree of correctness. The former component of probability the object exists is generally interpreted with the Average precision in the PR Curve and the latter is represented by IoU which can be labelled as a true object (Maxwell et al., 2021b).

3.7 Delineation of informal settlement

Generally, an illegal settlement has a distinctive characteristics and environment settings that enables researchers to delineate from the formal settlements. To delineate the informal settlement various researchers have identified indicators as depicted in Table 6 which includes widely used indicators but are not limited to:

Table 5. Informal settlement indicators as identified in various researches.

Sl.no	Indicators	Details	Authors
1	Built-up area	Residential dwelling size between less than 50sq.m and 380 sq.m as informal settlement	(Fallatah, 2020, Graesser et al., 2012)
2	Vegetation	Lack of incidental open space, play grounds and Vegetation mostly characterised by barren land	(Jain, 2007, Dare and Fraser, 2010, Fallatah, 2020, Graesser et al., 2012)
3	Lacunarity of housing structures	Spaces between built up areas in informal setting is not well planned and overcrowding. High built up density and orientation of built up.	(Fallatah, 2020, Graesser et al., 2012)
4	Road network	Irregular road network and not well planned without any setbacks from the built-up areas, characterised by numerous dead ends.	(Jain, 2007, Fallatah, 2020, Dare and Fraser, 2010, Graesser et al., 2012)
5	Texture and roofing material	Heterogeneity in use of roofing material like wood, plastics, concrete which have varying spectral properties	(Jain, 2007, Fallatah, 2020)
6	Location	Located in higher risk prone areas (Landfill sites, airport, transportation corridor areas) or in close proximity to city centers.	(Fallatah, 2020, Graesser et al., 2012)

3.7.1 Generalised Ontology description of informal settlement

In the Figure 30, it illustrates a general description of distinguishing properties of informal settlement from the formal settlement (Blaschke et al., 2008). The properties of informal settlement are not exhaustive as listed in Figure 30 , it can vary depending on the country specific environment, chronological land use change and other factors which cannot be generalized for the existence of informal settlement (Graesser et al., 2012). A settlement area is generally characterized with planned built up spaces, proper road network and accessibility, well designed open space and availability of vegetation in the areas but if the normally accepted settling has deviations in terms of physical properties its more likely to be grouped as informal settlement (Blaschke et al., 2008). But there are other legal, social and economic factor which can further validate such settings (Graesser et al., 2012).

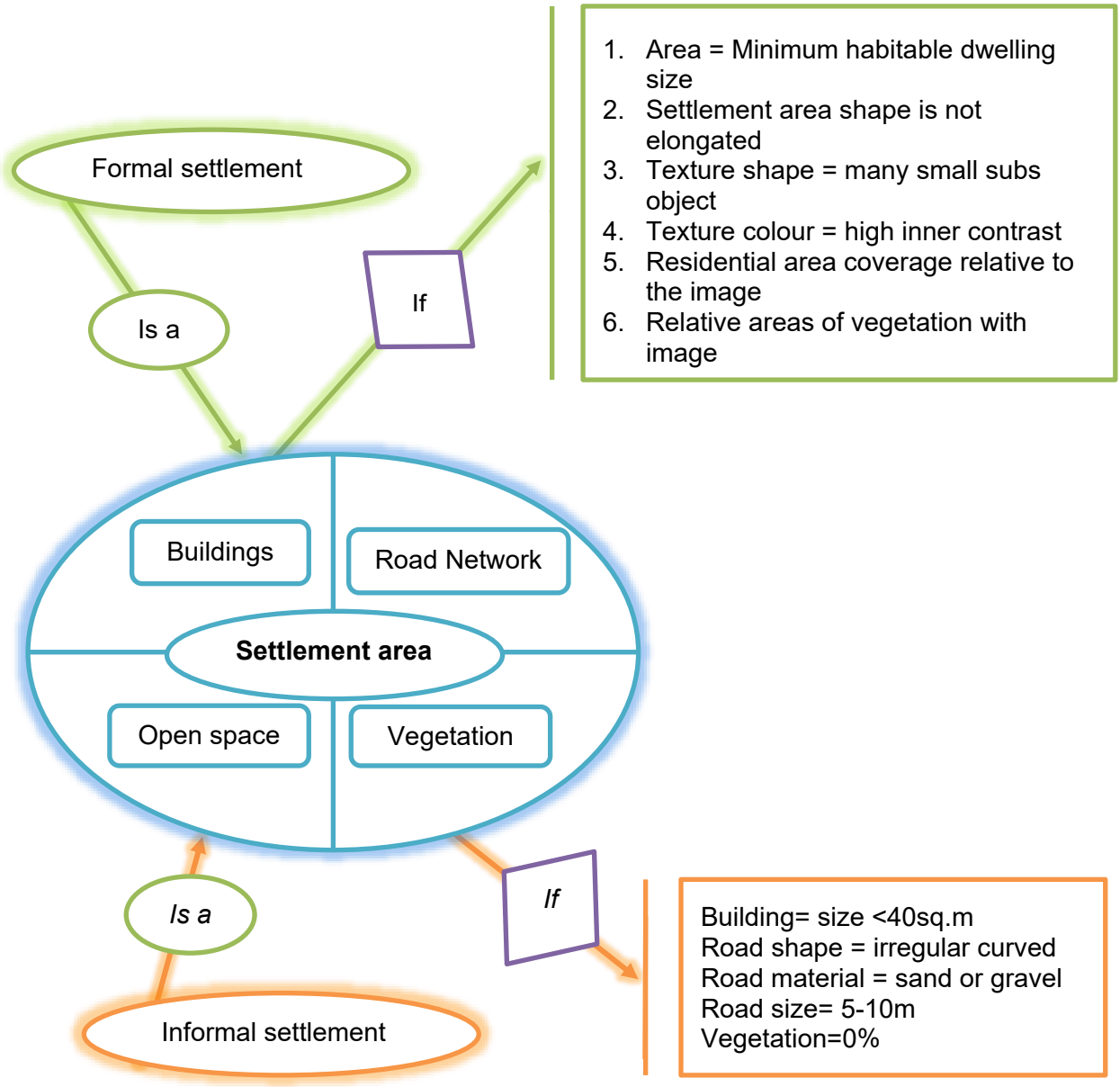


Figure 30. Ontology description of Informal and formal settlement

3.7.2 Ontology description of illegal settlements in study area

In Bhutan, the informal settlement is more related to the legality of the tenureship rather than the physical conditions of an informal settlement. There are still numerous land conflict and discrepancies related to land ownership. In the district where the study area is being conducted, there are total of 18, 932 land owners and the number of owners with disputed and conflicted land parcel are around 689 which are not yet resolved (NLC, 2012).

Bhutan is a mountainous country with extreme undulating terrain where only 0.16% out of the total arable land is under human settlement and approximately 70% is conserved as forested area (MoWHS, 2016). Such locational attributes and scarce land acts as an impetus to encroachments with growing urban population and rapid development. In Bhutan the cities are normally in a concentric and compact pattern and there are no concepts of suburbs. Settlement in rural areas is sparsely distributed and there are many parts of rural Bhutan that has no road connectivity and accessibility to basic infrastructures (MoWHS, 2016). To use generalized physical properties of informal settlement and apply on the study area won't be viable approach.

As presented in Figure 31, for this study area, Cadastral database for the district was collected. Upon overlaying the cadastral database with the classified image as produced from the OBIA approach and objects detected using deep learning approach, the illegal buildings can be identified.

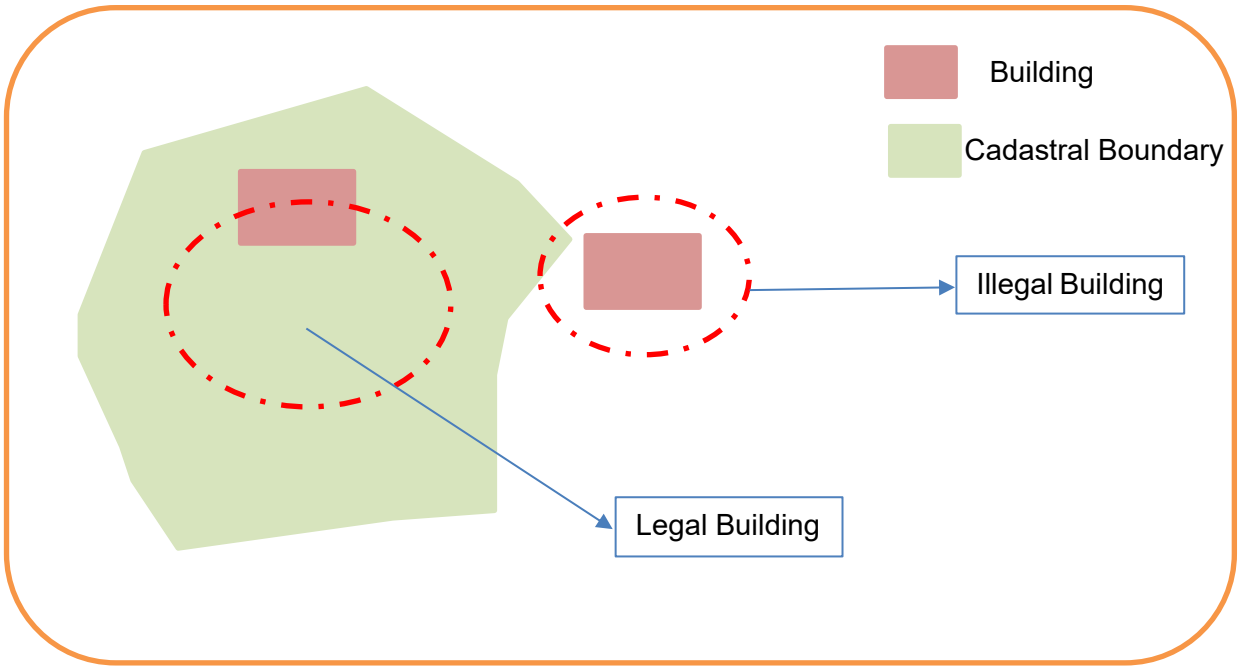


Figure 31. Illustrations showing the basic thumb rule for delineating illegal housing from legal buildings using cadastral boundary

4 CHAPTER: RESULTS

In this chapter the results that were achieved after carrying out the detailed methods as detailed out in chapter 3 for GeoEye-1 and Pleaides-1A imagery using Object based Image Analysis (OBIA) and Deep learning approach -Single Shot Multi-box detector (SSD) model respectively shall be presented. The results are bifurcated under the major analysis that were undertaken.

4.1 Orthorectification

In the orthorectification for Geoeye Imagery, the RMS error resulted from carrying out orthorectification taking reference dataset as the Digital Elevation Model was less than four pixels accuracy of Panchromatic imagery resolution. The RMS error for Panchromatic imagery using various transformation method is presented in the Table 7. The projective transformation was adopted for the orthorectification which is similar to rubber sheeting method was applied on images having significant errors. The RMS error value is a summation of all the residual error in x and y coordinate and the lower value not necessarily in all cases determine the level of alignment and rectification. Despite the high RMSE error with projective method, the image when overlayed and visually assessed, it seems to give a better alignment with base imagery than any other method. Then the orthorectified Panchromatic imagery was used for base reference to rectify the Multispectral imagery and the accuracy was less than three-pixel accuracy.

Table 6.RMSE using various transformation method for orthorectification of GeoEye-1 Imagery

Sl.NO	**Transformation method**	**RMSE (m)**
1	Zero Order Polynomial	106.99
2	1st Order Polynomial	41.14
3	2nd Order Polynomial	40.98
4	3rd Order Polynomial	39.37
5	Adjust	2.398
6	***Projective***	***41.01***

For the orthorectification of Pleaides-1A imagery approximately 646 control points were extracted from Panchromatic Imagery and 1st order polynomial transformation was used which

resulted in approximately 4m RMS error. Figure 32 depict the specification of control points, the Difference in X and Y coordinate and residuals.

Link	Source X	Source Y	X Map	YMap	Residual X	Residual Y	Residual
626	711,802.595037	2,973,810.937672	711,796.059337	2,973,814.974156	-1.177998	1.319729	1.769001
627	711,734.448221	2,973,708.962312	711,731.169806	2,973,712.471115	2.090172	0.785606	2.232935
628	711,740.205248	2,973,704.562997	711,736.276276	2,973,706.617357	1.439867	-0.668665	1.587555
629	711,724.761127	2,973,681.521046	711,719.391945	2,973,684.742336	0.002118	0.496807	0.496811
630	711,719.420729	2,973,679.248028	711,713.474219	2,973,682.038548	-0.574872	0.065696	0.578613
631	711,724.307068	2,973,668.646241	711,717.198305	2,973,671.376438	-1.736293	0.005322	1.736301
632	711,730.584254	2,973,671.758510	711,723.320091	2,973,674.641390	-1.892131	0.158418	1.898751
633	711,416.836572	2,973,958.631012	711,411.007001	2,973,960.732832	-0.475490	-0.630626	0.789798
634	711,426.215461	2,973,954.067894	711,419.424456	2,973,956.294537	-1.436749	-0.505450	1.523065
635	711,423.616515	2,973,947.946513	711,416.057474	2,973,949.815648	-2.204169	-0.863267	2.367191
636	711,372.466666	2,974,040.563516	711,372.891788	2,974,047.528707	5.772941	4.232801	7.158453
637	711,366.813592	2,974,041.723907	711,365.302141	2,974,047.917920	3.836406	3.461364	5.167112
638	711,365.640427	2,974,032.765976	711,365.691354	2,974,038.966029	5.399619	3.467087	6.416898
639	711,372.442684	2,974,032.566054	711,372.697181	2,974,037.992998	5.603037	2.694323	6.217186
640	711,388.974066	2,974,039.763179	711,387.876474	2,974,045.388038	4.249885	2.893298	5.141274
641	711,392.664602	2,974,039.190370	711,392.399125	2,974,045.388038	5.081959	3.466281	6.151538
642	711,392.905860	2,974,034.202660	711,392.274577	2,974,041.277952	4.716596	4.343775	6.412072
643	711,368.486058	2,974,049.140541	711,369.731381	2,974,056.597362	6.592470	4.724471	8.110566

Figure 32. Orthorectification of Panchromatic imagery for Pleaides 1A imagery

Geometric correction was conducted in ERDAS Imagine for Multispectral imagery taking Panchromatic imagery as reference dataset. Around 57 control points and 26 check points extracted for the polynomial geometric correction. The resultant check point RMS error are presented in the Figure 33. The geometric rectification resulted in less than 1 pixel accuracy of Panchromatic spatial resolution.

Point #	Point ID	>	Color	X Input	Y Input	>	Color	X Ref.	Y Ref.	Type	X Residual	Y Residual	RMS Error	Contrib.
56	GCP #56			712445.416	2972528.751			712445.978	2972528.064	Control				
57	GCP #57			712413.273	2972535.920			712413.810	2972534.273	Control				
58	GCP #58			708001.467	2975277.085			708002.900	2975275.908	Check	0.814	0.011	0.814	0.441
59	GCP #59			707983.375	2975258.171			707984.213	2975257.739	Check	0.225	0.752	0.785	0.425
60	GCP #60			707852.889	2975842.281			707853.127	2975841.007	Check	-0.448	-0.044	0.450	0.243
61	GCP #61			708082.907	2975905.982			708083.194	2975900.328	Check	-0.450	-4.388	4.411	2.387
62	GCP #62			708082.381	2975896.067			708082.081	2975894.035	Check	-1.037	-0.767	1.290	0.698
63	GCP #63			708130.599	2976012.759			708132.631	2976011.874	Check	1.267	0.399	1.328	0.719
64	GCP #64			708331.425	2976150.244			708333.232	2976149.634	Check	0.983	0.714	1.215	0.657
65	GCP #65			708308.159	2976174.058			708308.410	2976172.946	Check	-0.573	0.212	0.610	0.330
66	GCP #66			708273.346	2976386.121			708274.580	2976385.605	Check	0.381	0.826	0.910	0.492
67	GCP #67			708304.209	2976395.788			708304.046	2976395.162	Check	-1.023	0.720	1.251	0.677
68	GCP #68			708399.825	2976474.974			708400.751	2976474.613	Check	0.035	1.006	1.007	0.545
69	GCP #69			708442.058	2976943.935			708443.565	2976942.792	Check	0.530	0.280	0.599	0.324
70	GCP #70			708751.363	2977160.615			708749.458	2977160.897	Check	-2.975	1.767	3.460	1.872
71	GCP #71			708550.882	2977358.463			708552.049	2977357.249	Check	0.101	0.266	0.285	0.154
72	GCP #72			708375.614	2977361.425			708375.716	2977359.630	Check	-0.932	-0.336	0.991	0.536
73	GCP #73			708399.533	2977355.765			708401.440	2977354.947	Check	0.868	0.643	1.081	0.585
74	GCP #74			708350.964	2977351.382			708351.837	2977350.713	Check	-0.155	0.785	0.801	0.433
75	GCP #75			708118.796	2977835.127			708120.111	2977834.576	Check	0.248	0.926	0.959	0.519
76	GCP #76			706076.646	2977492.797			706078.710	2977490.755	Check	1.427	-0.858	1.666	0.901
77	GCP #77			706618.476	2976965.013			706620.906	2976964.230	Check	1.782	0.413	1.829	0.990
78	GCP #78			706993.642	2976729.822			706994.492	2976728.730	Check	0.173	0.125	0.214	0.116
79	GCP #79			707218.382	2976788.157			707216.146	2976792.000	Check	-2.964	5.095	5.895	3.190
80	GCP #80			707477.090	2976833.956			707477.991	2976832.943	Check	0.119	0.277	0.301	0.163
81	GCP #81			708115.876	2976192.651			708116.177	2976191.922	Check	-0.491	0.572	0.754	0.408
82	GCP #82			708111.198	2976284.687			708111.968	2976283.636	Check	-0.036	0.260	0.262	0.142
83	GCP #83			707986.646	2976007.869			707987.096	2976007.758	Check	-0.288	1.154	1.189	0.643

Figure 33. Geometric Correction for Multispectral Imagery for Pleiades-1A Imagery

4.2 Pan sharpening of raster images

Co-registration error between the Panchromatic and Multispectral imagery for GeoEye-1 imagery was very significant and the fused image wasn't effective for the desired processing therefore Multispectral was considered for further analysis. Figure 34, illustrates the discrepancies and noise data cause by overlapping of spectral reflectance in the fused image.

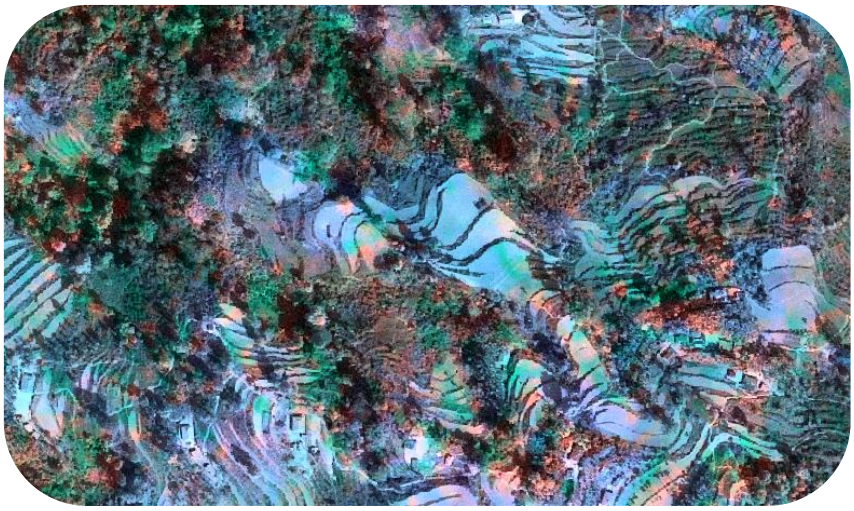


Figure 34. Pan sharpened image using GeoEye-1 imagery

In case of Pleaides-1A imagery enhanced resolution of Multispectral resolution from 2m to 0.5m was conducted using NN diffuse pan sharpening method is as illustrated in Figure 35.

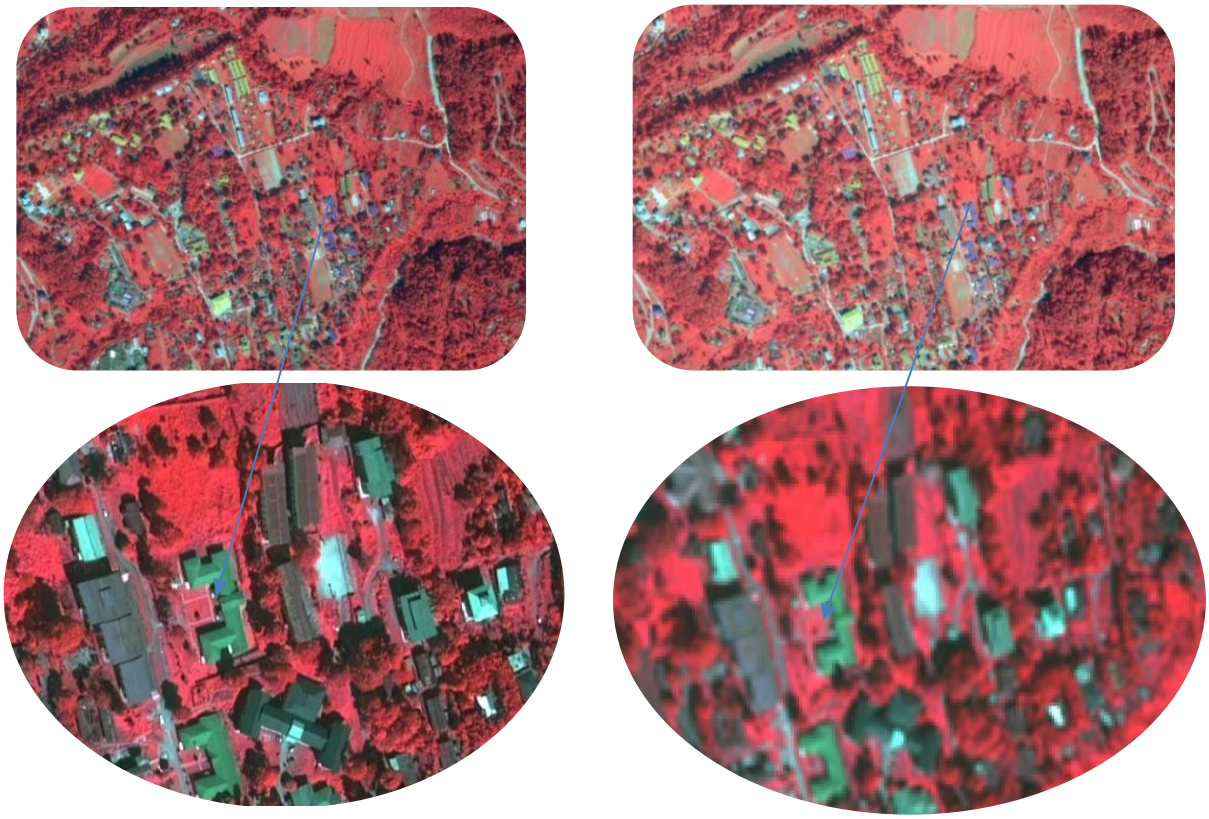


Figure 35. Left) Pan sharpened image, right) Raw Multispectral imagery

4.3 Object Classfication and Object detection - Resultant maps

As appended in Annexure B, the map represents the building classified using OBIA method and the total number of buildings detected was 682 in the selected tile. Annexure C, illustrates the building objects detected using Single Shot Multi-box detector Model and it has detected 2563 building objects.

4.4 Accuracy assessment

4.4.1 Accuracy assessment of classified image- GeoEye-1 Imagery

In this section it presents the results of accuracy assessment of the building features classified using OBIA approach based on the indicators and Quality metrices as discussed in section 3.6.1.

4.4.1.1 Result - Confusion Matrix and Kappa coefficient

The Table 8 shows the result from OBIA classification for four land classes using Confusion matric and kappa statistic to assess the accuracy and classification agreements.

Table 7. Confusion Matrix for the OBIA classification

Sl. No	**Class Value**	**Built-up**	**Road**	**Forest**	**Barren land/ Open space**	**Total**	**User Accuracy**	**Kappa**
1	Built-up	76	0	0	0	76	1	0
2	Road	0	52	0	0	52	1	0
3	Forest	0	0	191	0	191	1	0
4	Barren land/Open space	0	0	0	180	180	1	0
5	Total	76	52	191	180	499	0	0
6	Producer Accuracy	1	1	1	1	0	1	0
7	Kappa	0	0	0	0	0	0	1

The user accuracy depicts the false positives/ the omission error related to misclassification and in this case as the value is 1 as illustrated in Table 8, it represents the classification achieved 100% accuracy. The producer accuracy depicts false negatives or commission errors, and it resulted in value 1 (Table 8) resulting in 100% classification accuracy. The Kappa Coefficient is 1 which shows a perfect agreement of the classification and overall assessment (Figure 36)

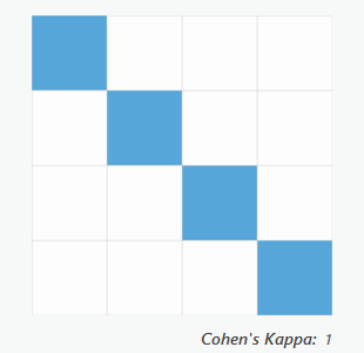


Figure 36. Kappa coefficient depicting perfect agreement for the Image classification

4.4.1.2 Reference data assessment and results

Before carrying out the assessment, it is essential step to validate and assess the reference dataset (Digitized buildings) that would be used to determine the quality assessment metrices.

In reference to section 3.6.1.2, the result from carrying out the analysis to validate the building database is as presented in Table 9 below:

Table 8. GIS analysis to validate the building database

Sl.no	Specification on analysis	Total number
1	Digitized building completely containing building point features	330
2	Digitized building containing building point within 5m buffer	*135*
3	Digitized building with no building point but there is existence of building when overlayed with the Multispectral imagery	*186*
4	Building point but not digitized	5 (some building points were not considered as the imagery was acquired before the building data was collected.
	Total Building digitized for the study area AOI	**656**

The building database thus acquired showed a lot of discrepancies and was not exhaustive but it was integral in validating the location of the digitized building. Thus, the digitized building was carefully constructed and taken up for accuracy assessment.

In the Venn diagram Figure 37, TP (True positive) the total number of buildings detected by both method and spatial analysis tool select by location was used to get the total figures.TN (True negative) is the total number of buildings that was digitized (reference dataset). FP (False positive) is the total number of buildings detected using semi-automatic approach (OBIA) and FN (False Negative) is the building left undetected by both approaches.

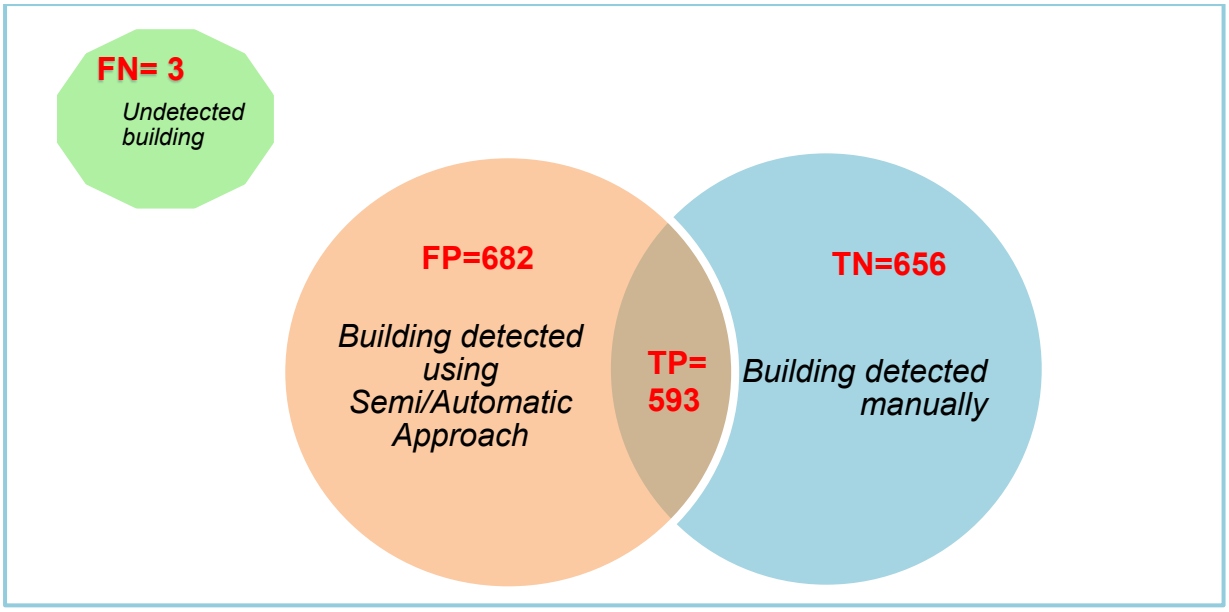


Figure 37. Venn Diagram to represent the result for quality metrices

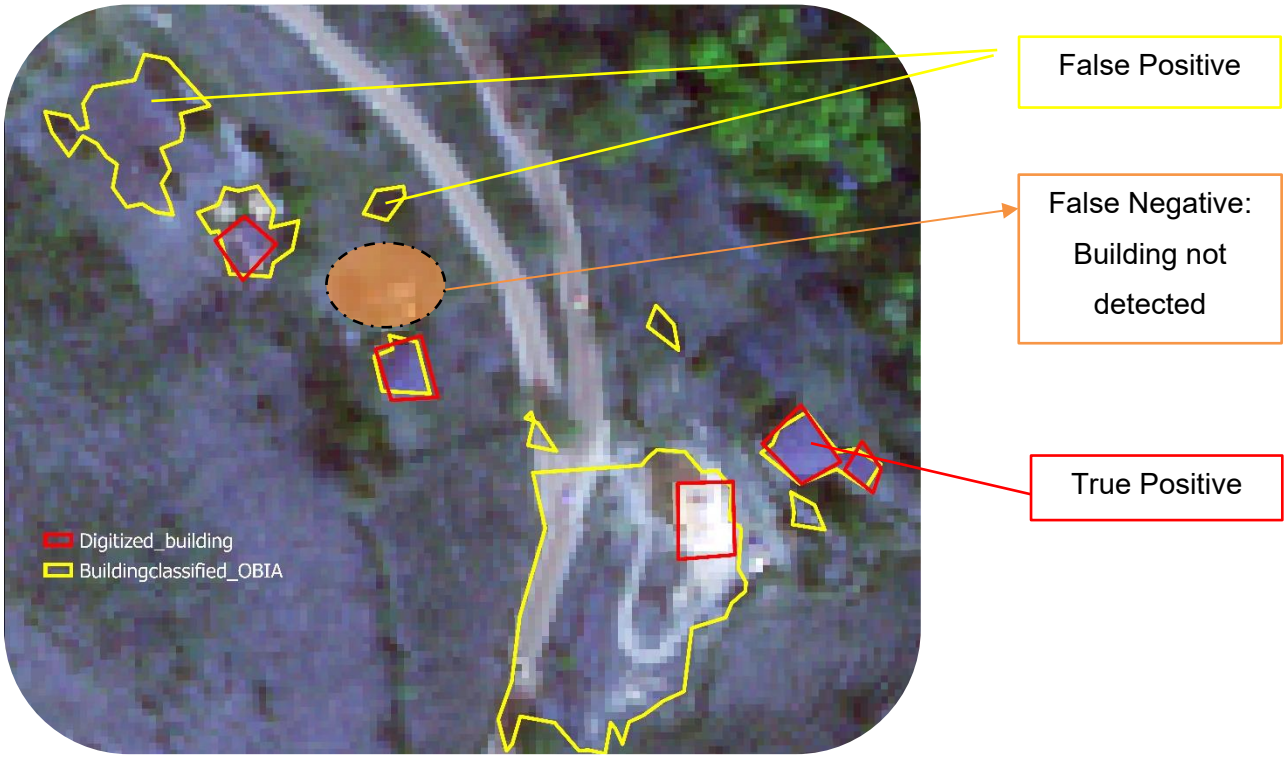


Figure 38. Illustration of True positive (TP), False Positive (FP), False Negative (FN) from the OBIA classified building data and manually digitized buildings

4.4.1.3 Area level Quality Assessment

As reported in Table 10, these metrics assesses the detection and performance of the Object classification method wherein the overall quality of the detection model is 46%, The branching

factor shows over detection with values 1.15, the omission error of 0.5% and detection percentage of 99%.

Table 9. Object level Assessment metrics

Value / Metrics	Branching factor	Miss Factor	Detection Percentage	Quality percentage
	1.15	0.51%	99.5%	46%

4.4.2 Result- Object level Quality Assessment

The correctness of object detected in the study area using OBIA approach is 51% and completeness percentage is 90.4%.

Table 10. Area level Assessment metrics

	OBIA classified building features	Reference dataset (Digitized building feature)
Correctly extracted object	350	593
Total Object	682	656
Assessment Metrics	Correctness = 51.5%	Completeness =90.4%

4.4.3 Accuracy Assessment of Object detected- Pleaides-1A Imagery

Accuracy of object detection model and the object detected using SSD model was assessed based on various indicators are presented in Table 12.

Table 11. Various indicators to assess the accuracy of the object detection model

Sl. No	IoU (%)	True Positive	False Positive	False Negative	Precision	Recall	F1 Score
1	IoU>=10	507	2058	1402	0.20	0.28	0.23
2	IoU>=20	272	18311	1661	0.015	0.14	0.027
3	IoU>=30	175	2390	1760	0.068	0.09	0.02
4	IoU>=40	87	2478	1849	0.03	0.04	0.039
5	IoU>=50	41	2524	1895	0.016	0.02	0.02
6	IoU>=75	1	2564	1935	0.0004	0.0005	0.0004
7	IoU>=95	-	-	-	-	-	-

The Figure 39, is the Precision plotted over recall indicators (PR Curve) which identifies the objects having IoU with 10 %. The mAP for IoU between 0.5 -0.95 is 0.000234.

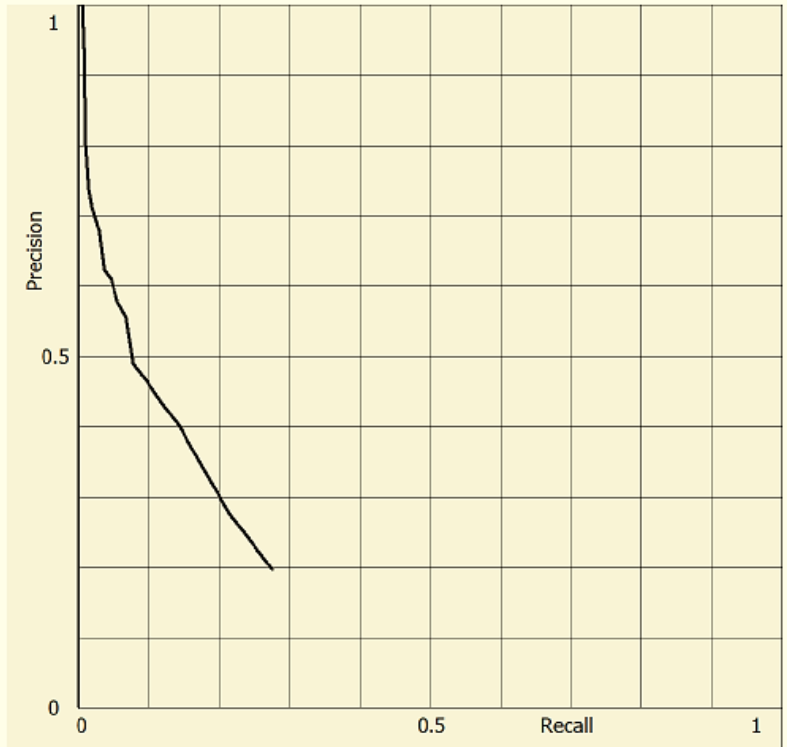


Figure 39 Precision-Recall (PR) Curve for IoU >=10%

4.5 Delineation of illegal settlement.

In Figure 40, the map shows the identification of illegal settlement after the image classification using OBIA approach. The techniques used for identification was overlaying the cadastral plot over the classified building map and the buildings without the legal parcel boundary was categorised as illegal buildings.16 buildings were identified using OBIA supplemented with cadastral information.

The bounding box generated after the object detection model was further processed by calculating the centroid of the polygon to generate the building point data. Taking the point data from the detection model, digitized building map was then overlaid with the cadastral plots to identify Illegal buildings as illustrated in Figure 41 the total number of buildings identified using the technique without parcel boundary was 48 buildings.

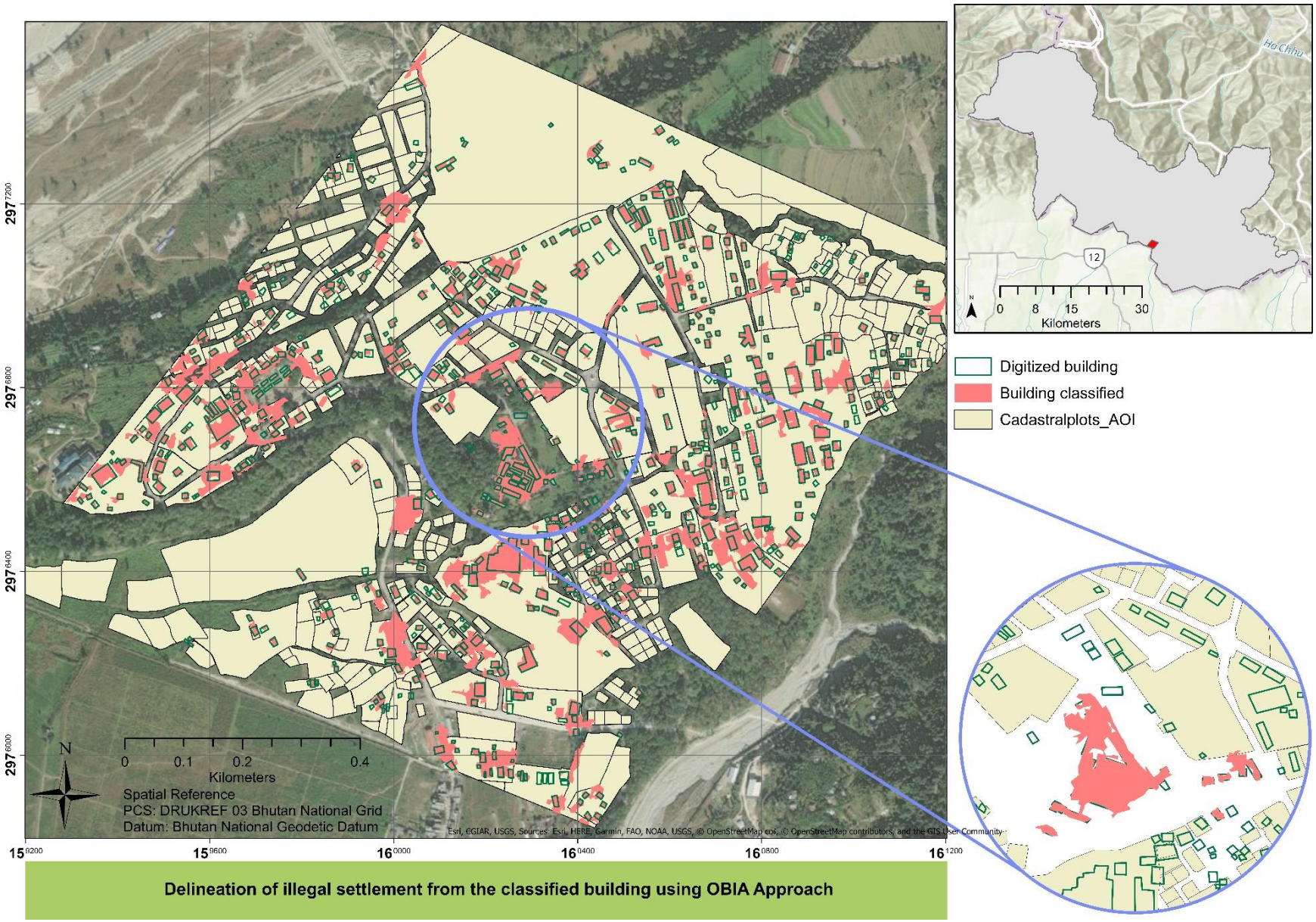


Figure 40. Illegal buildings identified using OBIA method and cadastral plots

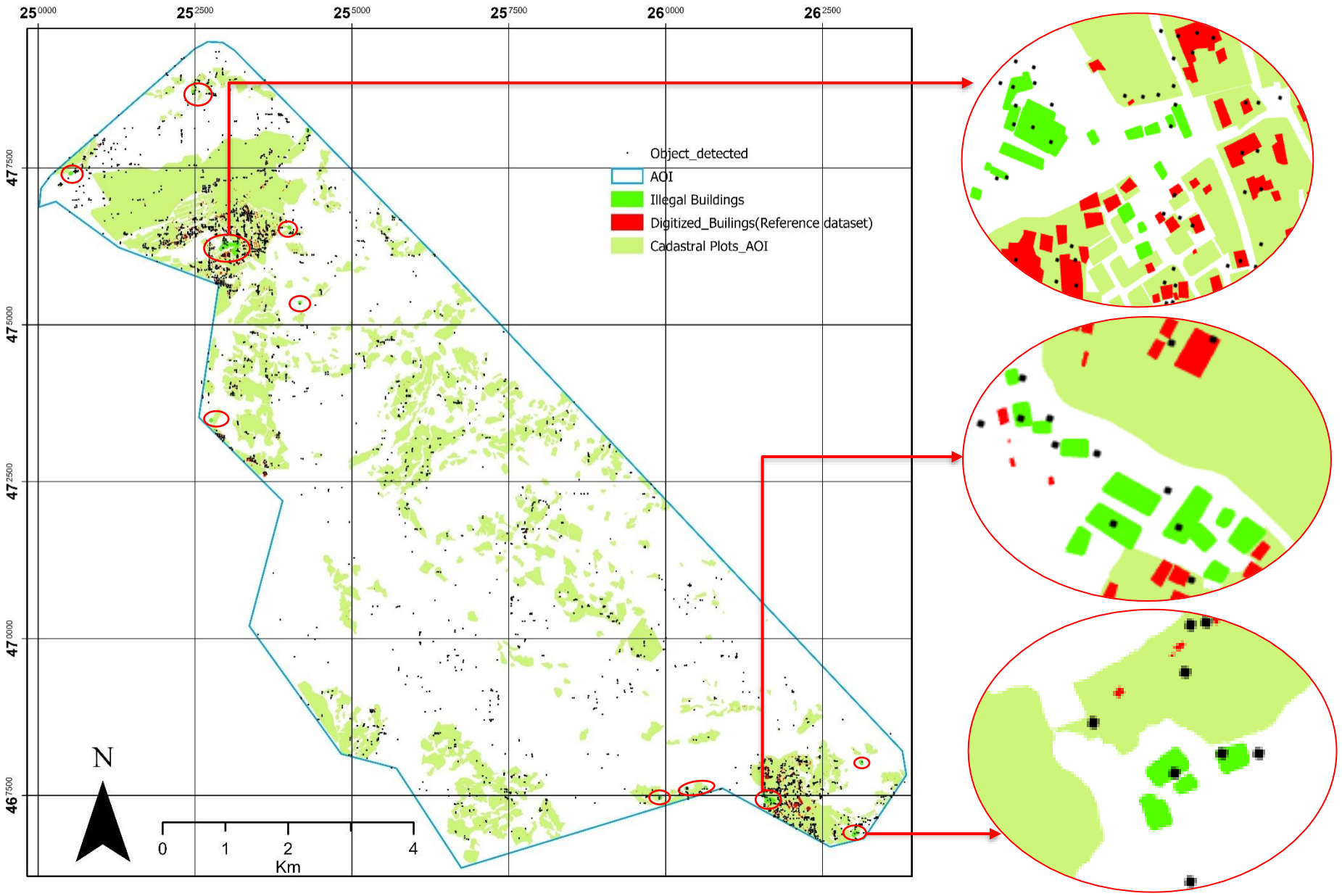


Figure 41. Illegal settlements identified using SSD Model assisted with cadastral plots and reference datasets

5 CHAPTER: DISCUSSION

In this Chapter, it shall present the key findings and observations from the methods and results as presented in chapter 3 and chapter 4. The research question as defined during the initial stages of the research streamlined the scope of the project and oriented the analysis accordingly. In this section it shall also investigate the extent the research questions were addressed.

5.1 Acqusition of imagery

Research Questions:

1. What is the minimum size of the dwelling units in the area of interest?
2. What is the minimum spatial and spectral resolution required to identify typical building structures in Bhutan?

The satellite imagery GeoEye-1 and Pleaides Imagery was acquired with spatial resolution less than the minimum size of the dwelling unit which was approximately around (5*4m). The very high resolution imagery like GeoEye-1 and Pleaides were used in many studies to conduct similar studies(Fallatah, 2020).

5.2 Image preprocessing

The GeoEye-1 Imagery that was acquired for the research significantly depicted co registration error between the Panchromatic and Multispectral imagery as well with the base imagery as available in the ArcGIS Pro. It is understandable that the accuracy of the google earth imagery available is also subjected to discussion on its locational accuracy (Aguilar et al., 2017) but nevertheless it serves as a base imagery for comparison and drawing inferences. The observation made above was presumed upon only visual inspection and to authenticate such inferences various orthorectification was carried out to rectify the significant issues. The first approach conducted was using Digital Elevation model to orthorectify the Panchromatic imagery but the process did not produce significant rectification. The root mean square error was less than four pixels but upon visual scrutinization, it was observed that the image needed more than a simple polynomial transformation. The deviation in the difference of the observed to the target location was not just one directional but it stretched out in various direction arising further discrepancies. Owing to the fact that region of interest has elevation ranging from more than 200 to less than 3500meter, it further makes the orthorectification more challenging. More than 600 control points were extracted and spline transformation was thus applied on the imagery. The spline transformation was applied and the RMS error was around 40m but upon

visual inspection it was realised that in some parts of areas the orthorectification using such method have performed considerably well while in some areas it generated more anomalies. After much consideration, the area of interest was further reduced and it included those areas where orthorectification performed well comparatively and the area of interest included different land uses which was essential for classification. The orthorectified Panchromatic imagery was used as the reference imagery to georeference the Multispectral imagery and the geometric correction on Multispectral imagery resulted in less than 3-pixel accuracy. But it was still a challenge to bring about considerable orthorectification for both imagery and such persistent error made the pan sharpening even more challenging as the fused imagery contained a lot of noise data. For Pleaides-1A Imagery, the orthorectification with simple polynomial transformation resulted in RMS error 8-pixel accuracy (considering the spatial resolution of Panchromatic imagery) and less than 2-pixel accuracy for Multispectral imagery. The pan sharpening process was conducted and the resolution of the imagery was enhanced to 0.5m as that of Panchromatic imagery. The Nearest neighbour Diffusion method was adopted in similar studies for spatial resolution enhancement and without significant distortion in the spectral and spatial properties (Sun et al., 2014)

5.3 Image analysis- Image classification and object detection

Research Question: What are the most appropriate methods for detecting buildings in high resolution satellite imagery?

The object-based image analysis is widely adopted for many fields of interest and for this project as well the approach selected was OBIA approach. Traditionally pixel based was generally used for image classification but with changing advancement, satellite imagery is now available with increased spatial resolution hence the pixel level approach does not provide sufficient information for classification and moreover the approach doesn't compliment with other geospatial tool for further analysis (Nussbaum and Menz, 2008). The degree of classification and accuracy is dependent on three primary steps in OBIA:

Image segmentation: it is a supervised and iterative process that entails selecting suitable parameters for creating meaningful segment based on the spectral and spatial significance. Much focus was given on segmenting the building features and the parameters setting was encircled to detect and segment individual buildings but also to detect the smallest of building units. Segmentation was based on the principle that most of the heterogenous features were consciously demarcated but refraining further bifurcation of homogeneous regions into sub-segments and avoid over segmentation (Nussbaum and Menz, 2008).

a. **Collecting training samples:** based on the land uses available on the area of interest the training samples were collected for all land classes defined. There was an attempt to classify the imagery based on different colour of roof types including other land classes but such process didn't produce satisfactory result as there were limited samples under some classes which led to misclassification. So, the features irrespective of roof colour types were categorised as one class because the main aim of this research is limited to detecting building footprints and not the study of the spectral behaviour of roof types and how it affects the classification. Another attempt was made with limited training samples for classification and it resulted in a lot of misclassifications of buildings with the bare soil and unpaved roads. Such attempts further validated the importance of having ample training samples which are representative of all the features present in the area with respect to size and shape to enormously reduce the overfitting issue with SVM classifiers (Norman et al., 2021). The shape of training samples also has direct impact on classification, especially for building with well-defined edges and shape which considers the rectangularity parameter for classification (Norman et al., 2021).

b. **Image classification:** supervised object-based classification was selected for this research and Support Vector machine classifier was used for the classification using Multispectral imagery. In Figure 42, it illustrates the results from the image classification.

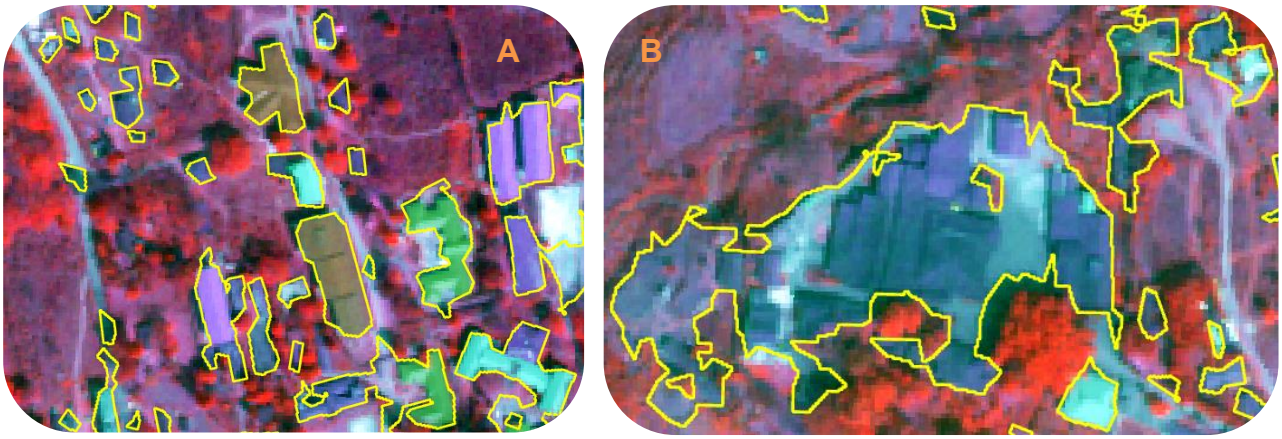


Figure 42. A) Image classification showing well classified Buildings B) generalised boundary for building classified.

In the Figure 42 A) is an example of classified image showing better result of building classification. In this instance, the smallest of building units are classified by the approach but the boundary definition is not accurate but for the project per se it has remarkably classified building classes. But in the Figure 42 B) it has generalised the boundary containing numerous building which could have been attributed by the similar spectral reflectance from the objects in the close vicinity and thus classified as one object. But even with such classification it is

very clear that there is inclusion of all the buildings inside the generalised boundary. For such classification a post processing is necessary to segment into individual features.

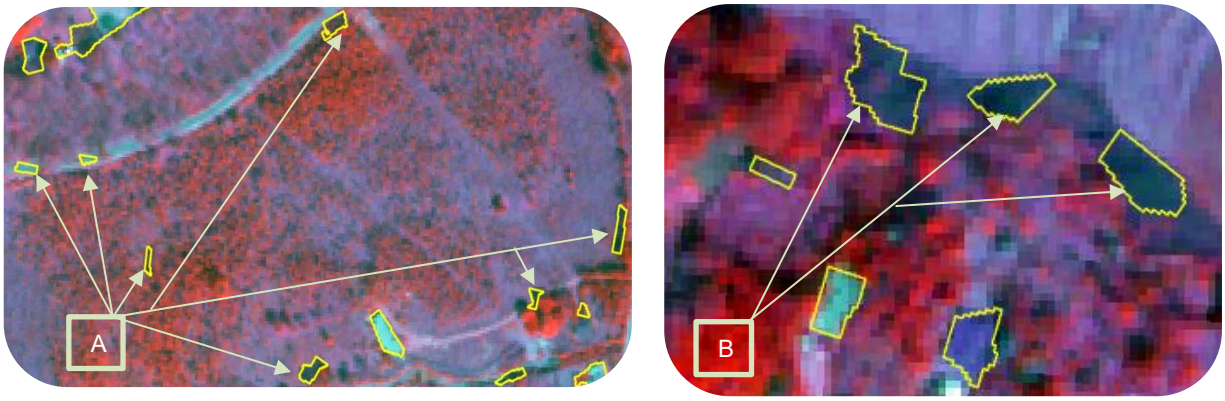


Figure 43. Misclassification in other land classes A) Road classes and shadow classified as buildings B) Barren land with darker spectral reflectance misclassified as building.

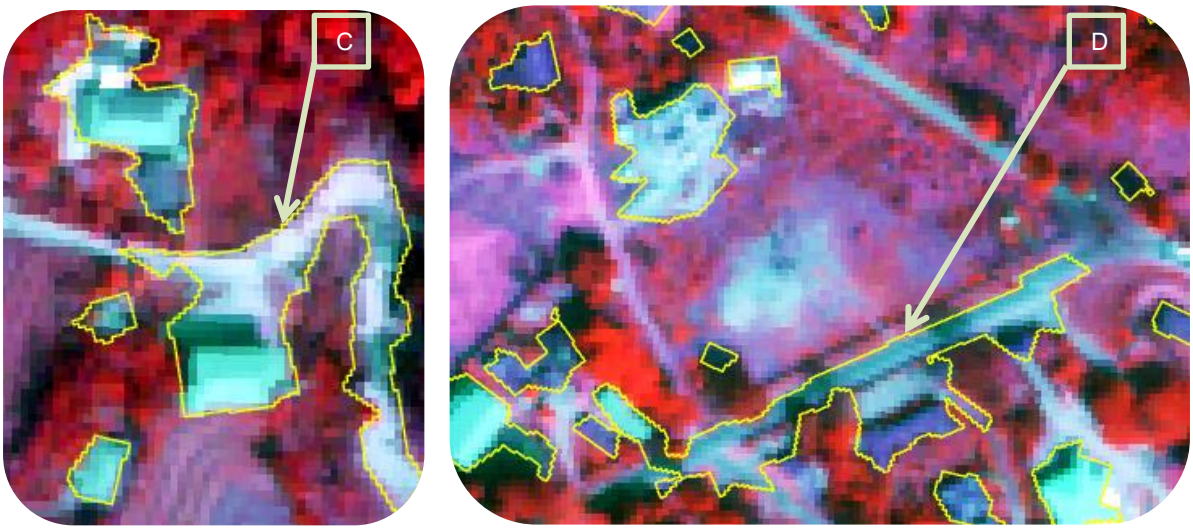


Figure 44. C&D) Road features misclassified and generalised within building boundary

In this Figure 43 and 44, it depicts the misclassification instances. There were numerous spectral overlaps of rooftops with the bare soil and road. Such discrepancies have led to misclassification of buildings in road or bare soil classes. In southern part of Bhutan there are varying roof types made from asbestos, tin sheet, slates, concrete tiles, rammed earth roof etc and some roads are concrete based, some are unpaved (sand, gravel, bare soil) and some paved roads. Such variability in materials and spectral similarity must have attributed to misclassification.

Another suitable approach considered for the project was deep learning approach which in recent years have been gaining a lot of attention because of its increased flexibility and incorporation of remote sensing data and Geospatial analysis infused with computer vision (Hoeser and Kuenzer, 2020). Contemporarily, the state-of-the-art object detection models

have variability in the incorporation of components: bounding boxes and the classifiers adopted and resampling of pixel/features (Liu et al., 2016). PASCAL Visual Object classes (VOC) is one of the detection yardsticks that are adopted in other models like Faster R-CNN; this model is considered to provide accurate detection results however it is demands intensive computational processing time and technically sophisticated hardware which is deemed not viable for real time applications (Hoeser et al., 2020). The speed of Deep learning model is generally expressed in frames per seconds; Faster R-CNN which is considered as one of the fastest amongst R-CNN Models, providing high accuracy but it processes only 7 frames per seconds, YOLO model processes 45 frames per second. But on the other hand, SSD model operates at a much faster pace with about 46 frames per second and it dismisses the stages of resampling the pixel by creating layers at different scales to detect objects yet generating higher accuracy than other models (Liu et al., 2016, Vaidya and Paunwala, 2019).

In ArcGIS Pro, it enables the users to carry out the deep learning processes and presents with numerous object detection models. The model adopted for the research is Single Shot Multi-box detector Model using ResNet-34. It is an end-to-end training producing higher accuracy and moreover considered as an optimal solution in object detection contemporarily (Liu et al., 2016, Vaidya and Paunwala, 2019). The integral process of deep learning approach is training the samples and the accuracy of the object detection has direct relationship with the training samples and the comprehensive representation of objects while training the model (Zhao et al., 2018). The model incorporates the training samples and then generates the prediction based on the samples as depicted in Figure 45.

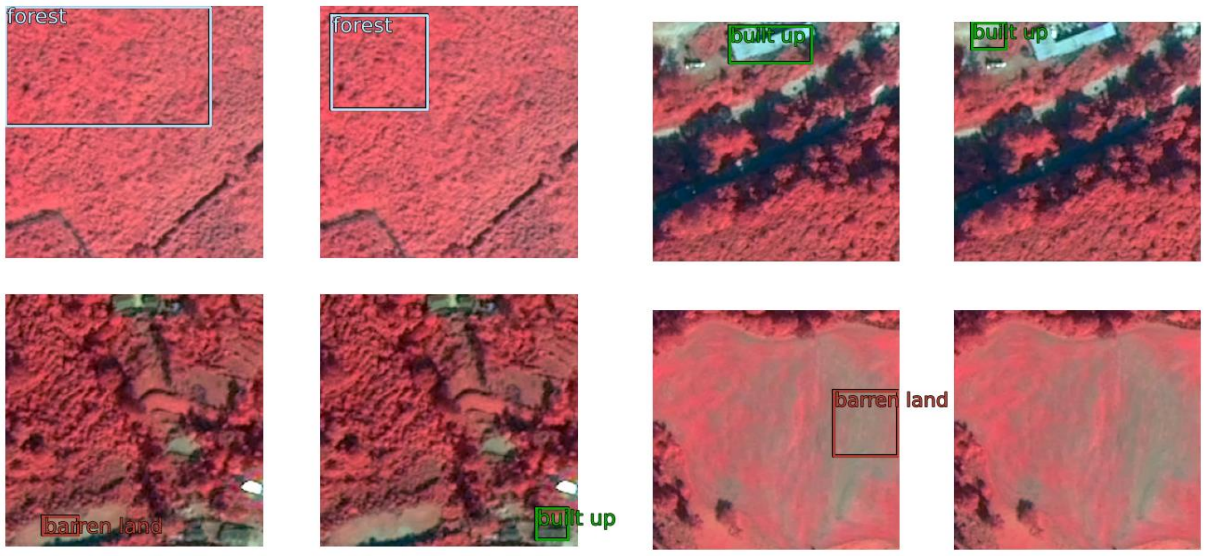


Figure 45. Prediction of classes after training the model

The training data are used for training the model helps in building the weights and filter for the network-based detection. 50 epoch was chosen for the study area which implies the number of iterations run over the entire training dataset. Another set of data called validation image dataset is compartmentalised and every after one epoch the performance of the model is tested on the validation dataset. 20% of the data was used as validation dataset for the project. The output generate from above process is training -validation loss graph which helps the user to understand the need for additional epoch and fitness of model; overfitted model has the decreasing training loss with increasing training samples and the validation loss increases with increasing training samples causing overfitting situation leading to generalisation of samples causing misclassification and thus more samples must be included to rectify (Salman and Liu, 2019). The reverse is the case of under fitted curve where the training loss increases with training samples but at certain point reaches an arbitrary point which implies even with further training the performance of the model won't be affected (Liu et al., 2016, Salman and Liu, 2019). For the study area the curve resembles to that of a good fit curve as illustrated in Figure 46 where the validation and training loss curve is close and validation curve placed a little higher. Validation loss curve shows a decreasing trend initially but after certain point takes a flattened curve which depicts that the increasing training doesn't improve the model after certain extent (Salman and Liu, 2019).

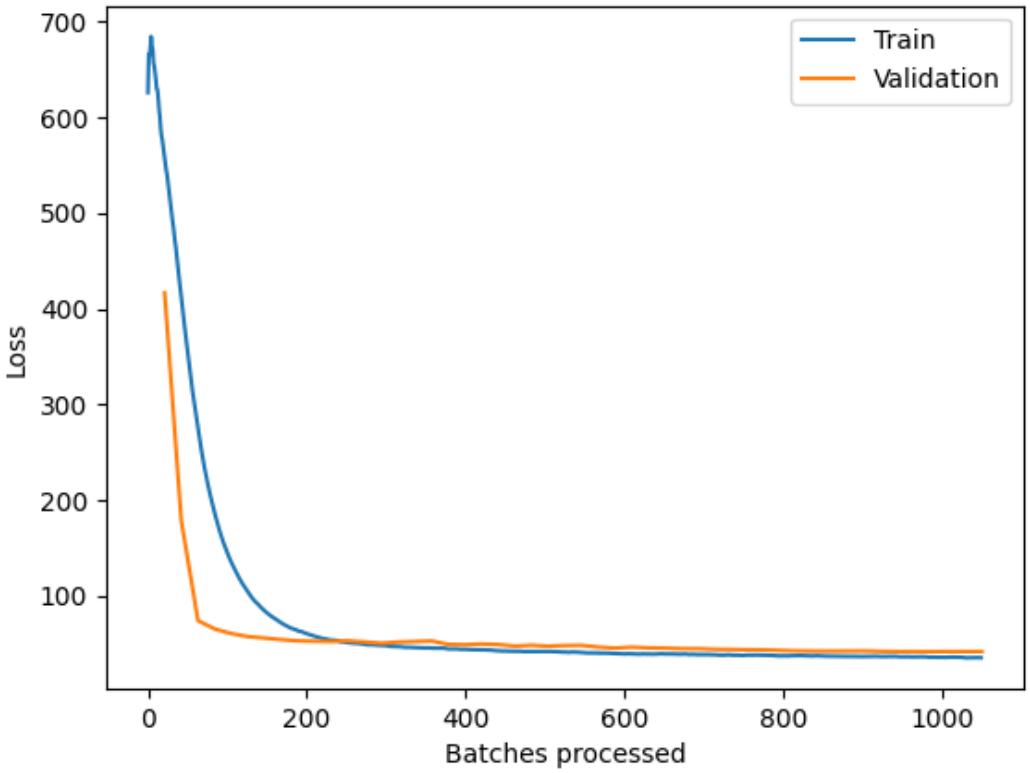


Figure 46.Training and validation loss curve to depict the fitness of the model

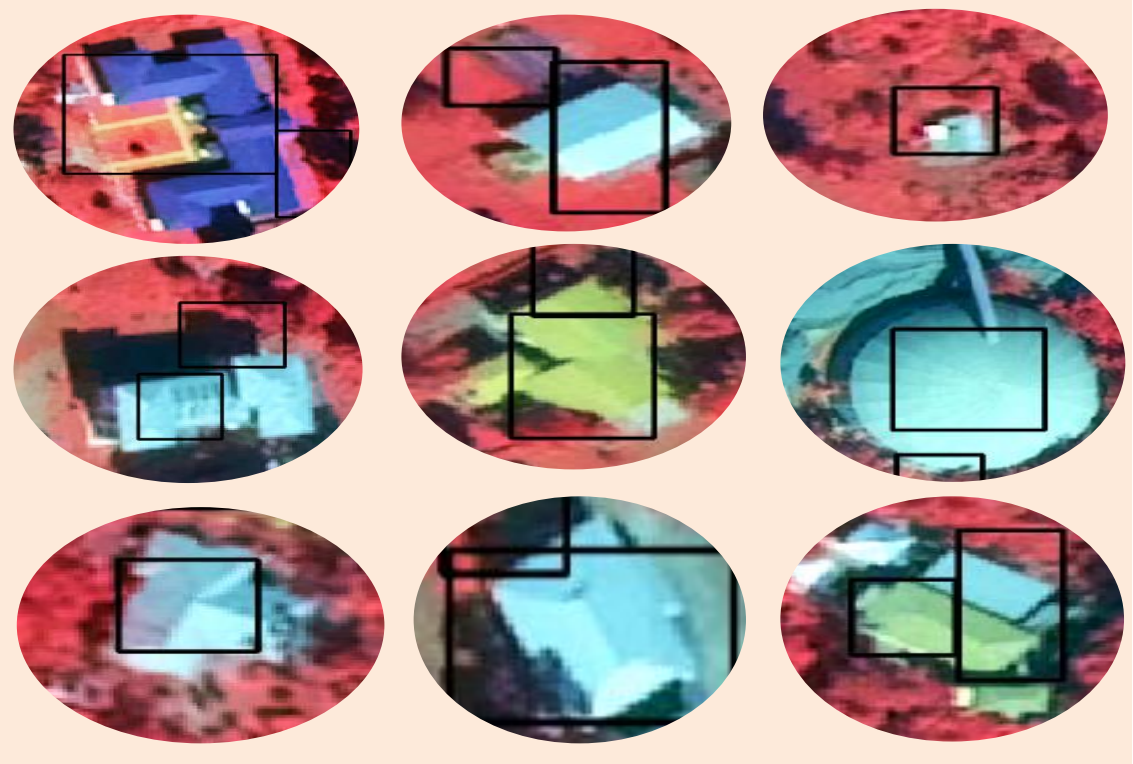


Figure 47.Illustration of buildings detected correctly and misclassification using the SSD Model

In Figure 47, are the extracts from the objects detected using SSD model where the building objects have been correctly detected. Over 500 samples were collected and incorporated for training the model and samples of building with varying sizes and shapes were also trained to ensure a wholesome representation. Other prominent land classes like road, barren land, river and forest were trained to assist in the detection of objects. After the object detection were complete, only the buildings layer was exported which were then taken up for accuracy assessment.

In Figure 48, it depicts the misclassification of building objects on the paddy fields, road surfaces and the river banks. The model has generated many false positives for building objects.

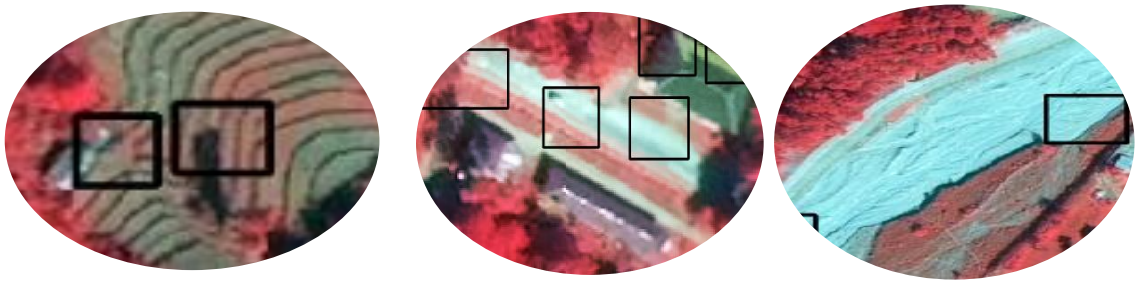


Figure 48.Incorrectly detected building objects on other land classes like paddy fields, road and river areas.

5.4 Accuracy assessment

Research Question: What are the most appropriate techniques for accuracy assessment of the data generated using various methods?

In this research, confusion matrix and Kappa coefficient was adopted which is a traditional accuracy assessment for remote sensing data. Such technique is based on the principle that the points generated for accuracy assessment from the image are random and unbiased that doesn't overlap with the training samples as it is assumed that the samples inherent uncertainties associated (Maxwell et al., 2021a). The accuracy assessment conducted for most of the OBIA approaches are by using kappa coefficient and confusion matrix to assess the classified image (Fallatah, 2020, Norman et al., 2021, Blaschke, 2010). For this research, the kappa coefficient value was 1 which depicts perfect agreement of the classification and accuracy of the output from the classification. It is still an active discussion on the units of assessment to be pixels, objects/ features, polygon or a group of pixels (Maxwell et al., 2021b). It is considered that Kappa is indeed a traditional and not a suitable indicator of accuracy assessment but rather it complements the adjustments done for overall accuracy. The kappa is strongly influenced with the class representations and the imbalance greatly affects the accuracy of the classification (Maxwell et al., 2021a). Therefore, another approach of accuracy assessment was resorted which involves the manual digitization of building from the imagery (Gavankar and Ghosh, 2018, Benarchid et al., 2013). The limiting factor for such approach is the scale of study area and settlement coverage. The digitized building polygon served as a ground truth and assisted in carrying out object and area-based metrics to generate the quality of classification, performance of the detection model and misclassification percentages. The OBIA model performance was reported as 46% which was attributed because of the detection percentage of 99.5% showcasing numerous false positives because of misclassification. The most prominent factor that led to misclassification was the overlapping of the bare land, unpaved road emitting similar reflectance as that of some bright roof tops. The branching factor (1.15) was over 1 which depicts the over/misclassification of objects. But for the overall classification accuracy based on the digitized polygon, the completeness percentages were reported as 90% which shows that OBIA based classified image have classified 90% of the buildings correctly.

As with the deep learning approach also, SSD model requires training samples for model training and the trained model has information with regard to the spatial context of image data. The data used for training and detection is not a pixels or group of pixels forming object like in case of OBIA but rather the input data is the subset of image (Image Chip) which is user definable (Maxwell et al., 2021a). OBIA is perceived as a semantic segmentation and

classification which includes identification of multiple classes and land uses. But Object detection is an instance segmentation which may include multiple classes or just single class. Use of confusion matrix is considered not viable for the accuracy assessment of Deep learning objects as the true negative detected through the model cannot be defined in a specified class and it nullifies the concept of using confusion matrix. The other metrices used for the deep learning approach as well as for the project are reported in the section 3.6.2. Even for this research, manual digitization was considered to build the reference datasets in computing the metrices for accuracy assessment.

The results from the object detection model with varying IoU tested for the study area depicts that most of the objects detected has overlap percentage of 10% and 507 building were deemed as building in reference to the ground truth data. The model performance was 20% with Pascal VOC dataset. The higher value for IoU led to increase in the misclassification or the objects that were left undetected in the study area. The Precision -Recall curve in Figure 37 suggest that the performance of the model is not good and accurate, as initially the precision was high but with increasing training samples and recall, the precision steadily dropped and diminishing after IoU>=75. Several factors could have impacted the lower accuracy of the object detection:

a. Unequal representation of training samples: the four land classes chosen for the study area were represented more by building samples. Several attempts were made before finalizing the model output for the study area; one such attempt which identified buildings much better were by limiting the number of classes but the drawback of such attempt was misclassification owing to the lesser training samples and only two classes. As the approach is object detection and not the semantic image classification which needs representation of land classes, limiting the class and increasing the quantify of the samples would have helped the accuracy. The loss and validation graph from Figure 44 also suggested that the training samples after certain point didn't not help improve the performance because it requires correct representation and considerably more samples with regard to the study area extent.

b. The SSD model does have the option of *Non maximum suppression* which removes the replication of the bounding box overlapping the features but it doesn't fully eliminate such error. The overlapping bounding box is either a true positive depending on the overlap percentage or a false positive which in turn account for the computation of the accuracy. Another similar situation is when the object detected over the feature has overlap percentage less than 50% which is normally considered for most of the studies, has the drawback of ignoring detection with IoU falling below. The location of the bounding box is not considered but rather it computes all the accuracy metrices based on the area of overlap which creates a

lot of anomalies. In the Figure 49, The bounding box has captured the location of the building correctly but the union of the building with the reference data is more than the intersection portion resulting in lower IoU. Lower IoU results in more false positives and misclassifications.

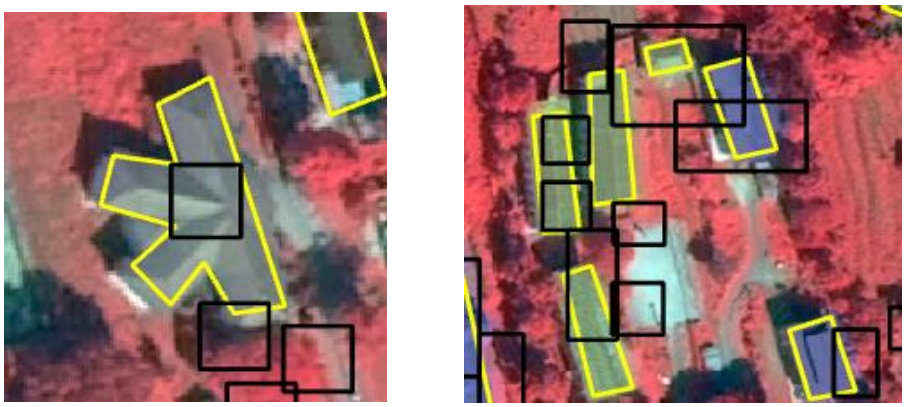


Figure 49. Overlapping bounding box for the object detected

c. The SSD model outperforms other model in terms of processing time but the model has poor performance in detecting smaller objects (Cao et al., 2020).

Bounding box generated from the detection model doesn't represent the building outline but it provides the location of the building which can be approximated into a point data format. As the aim of the project was not to trace the building footprints but rather validate the location of the building and then identify the illegal settlement.

For both the approaches post processing which shall include, rectification for the misclassified objects/features and reclassifying those objects by referencing to ground truth data and regularizing the building can also be conducted to further sieve the anomalies and create a clean classified image or object map. For this research, post processing is not in the scope of this research as the objective was limited to identifying the appropriate method and reporting model performance and accuracy of the classifications.

5.5 Delineation of illegal settlement

Research question: What are the likely environments to identify informal settlements in Bhutan?

Informal settlement in Bhutan is more associated with the illegality and discrepancies in the tenureship rather than the physical conditions contemporarily. But such situation may become more of informal settlement in long run if not assessed and monitored in the early stages. Therefore, instead assessing the likely physical environments for identifying illegal settlements, the land parcel boundary was used for delineating the illegal plots form the legal plots for this research. Using OBIA approach with cadastral information for the small AOI, 16 building was were thus identified as illegal buildings. Using deep learning model, even though the accuracy assessment generated poor model performance, the location of the bounding box was considered as the building point and 48 buildings were identified as illegal Buildings. Adoption of such methodology and upscaling the project area would be very beneficial to the government in identifying the illegal plots and then ground truthing can be thus carried out based on the findings from such studies.

6 CHAPTER: CONCLUSION AND RECOMMENDATIONS

Object based image analysis and deep learning indeed offers a promising potential in feature classification and object detection respectively and most importantly integration of various data sources for spatial analysis. It was inferred from the research that the most integral procedures that determine the accuracy of classification and detection is the segmentation performance and representative yet ample training samples (Nussbaum and Menz, 2008). The output produced from the OBIA method and Deep learning approach became a major input for identification of illegal buildings in the study area. The challenge faced in this research while conducting the Object based Image classification was related to the co-registration of the imagery which could have been rightly addressed with extensive pre-processing methods and high-resolution Digital elevation Model or ancillary data that compliments the pre-processing process. The time bound research expedited the process using more suitable processing and appropriate alternatives in the interest of project and time. Nevertheless, using the geospatial analysis and output from OBIA classification assisted in identification of illegal settlement in the smaller AOI. Considering the scope and aim of research another very high-resolution Imagery Pleaides-1A was acquired on 09/10/21 to test the deep learning algorithms and outline the differing methods involved for both the approach but the main objective encircles around identification of illegal settlement in the defined area of interest. The OBIA approach resulted in higher accuracy whereas the deep learning approach couldn't yield the desired accuracy but the key findings from the approach can be an important backbone for further analysis in future. For both the approach, level of accuracy was directly dependent on the comprehensiveness of the training samples and the selection of parameters which are area specific. The research was aimed at building a basic methodology and presenting with alternative approaches that can be adopted in identification of illegal settlements. Such methodology will not only assist in the identification of illegal buildings and the upscaled for entire Bhutan but also shall be a major recommendation and input for rehabilitation program for the landless and vulnerable section of people.

6.1 Recommendation and limitations

The OBIA approach was conducted in ArcGIS Pro software which had very less room for flexibility and inclusion of other parameters. In the field of remote sensing since 1999, eCognition software have been widely adopted for object-based image analysis. Several studies conducted to assess the classification accuracy using high resolution satellite imagery provided by various other softwares (Data Dissection tool, ERDAS Imagine, CAESAR, InfoPACK) further validate that eCognition outperformed other softwares (Meinel and Neubert, 2004, Fallatah, 2020). As eCognition meets the demand of multiscale segmentation based on fuzzy logic (Li et al., 2005, Nussbaum and Menz, 2008) and integration of data with existing remote sensing and geospatial programs (Nussbaum and Menz, 2008).

As a major finding from the OBIA approach, it was inferred that road and roof classes showed similar spectral characteristics and such overlap hampered the overall accuracy of the classification, hence such limitations could be tackled by use of photogrammetry and mapping using UAV (Veeranampalayam Sivakumar et al., 2020). UAV which provides high resolution imagery and generation of digital models which can become a major input for the orthorectification along with the use of primary ground control points. The methods are cost effective and flexible solution in acquiring high resolution imagery but the drawback in using such technology is the spatial extent for flights and the processing factor (Gevaert et al., 2017).

The major challenge for object detection is the real time application as the model given the capacity to incorporate large dataset it requires longer computation time and sophisticated hardware in terms of the GPU and CPU (Cao et al., 2020). The high-end computers also require considerable number of days for just training the model (Maxwell et al., 2021a). For this project given the limitation of time and high-end hardware requirement, a more viable alternative which produces high accuracy with less computation time was adopted. SSD model outperforms other model in terms of processing time and it has a simpler training modules but the model often tends to neglect the smaller objects in the image and requires ample training samples for computation (Cao et al., 2020, Park and Kim, 2020). But various studies have attempted to improve the model by incorporating multiple datasets while training the model which can improve the detection rather than just relying on the image chip generated from one dataset (Cao et al., 2020). Furthermore, the trained model should be assessed on different imagery with varying environment than the study area to better understand the robustness of the trained model and outline the key findings to further improve the model upon investigation.

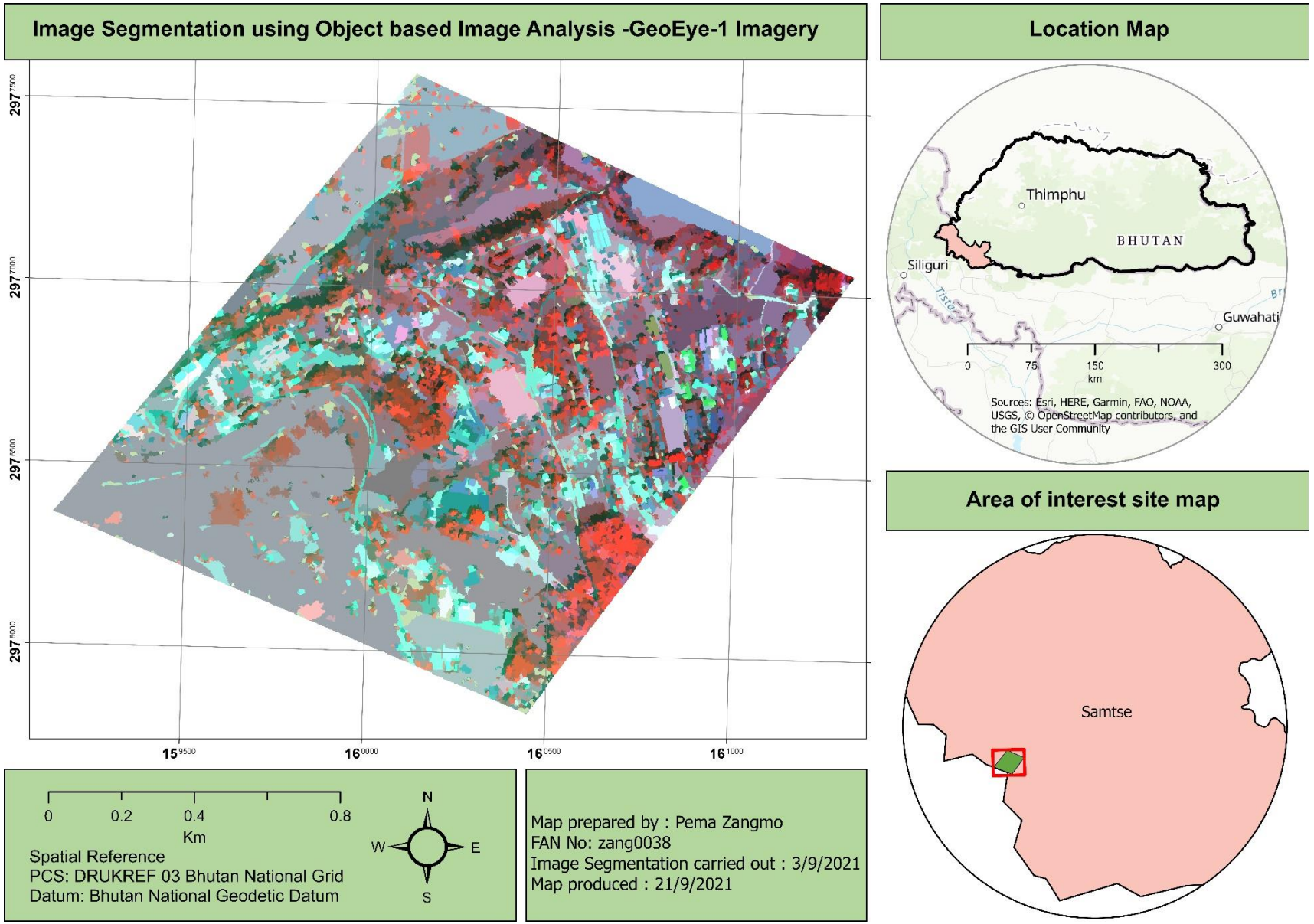
The satellite imagery consists of multiple bands other than RGB bands but for the data to be used for the deep learning model the data are normally processed in the Three band (RGB) combination hence the object recognition and detection could be more challenging even if learning models have been attuned with image compression for better computation. Most of the studies have presented the most challenging part of the deep learning was training the model from the base and due to lack of data it could generate uncertainties when the users have limited idea on chemistry and mechanism of different parameters and the interrelationship of various algorithms in a deep neural network. Transfer learning have been recommended by various researchers as one of the decentralised techniques that presents with already pre-trained model and best practices on a larger dataset and the users can adopt for smaller area of interest with inclusion of area specific parameters and modelling (Maxwell et al., 2021b, Veeranampalayam Sivakumar et al., 2020). Such adoption has resulted in remarkable model performance (Veeranampalayam Sivakumar et al., 2020, Hoeser et al., 2020).

Another approach that can further accentuate the deep learning model performance in using high resolution imagery from UAV. To expedite the processing time which was a concern for using such technology was managed by not mosaicking the image extracted from the UAV but rather processing it separately and refraining from the overlapping error (Veeranampalayam Sivakumar et al., 2020). The satellite imagery owing to the pre-processing processes might result in noise data that can hamper the object detection Veeranampalayam Sivakumar et al. (2020).

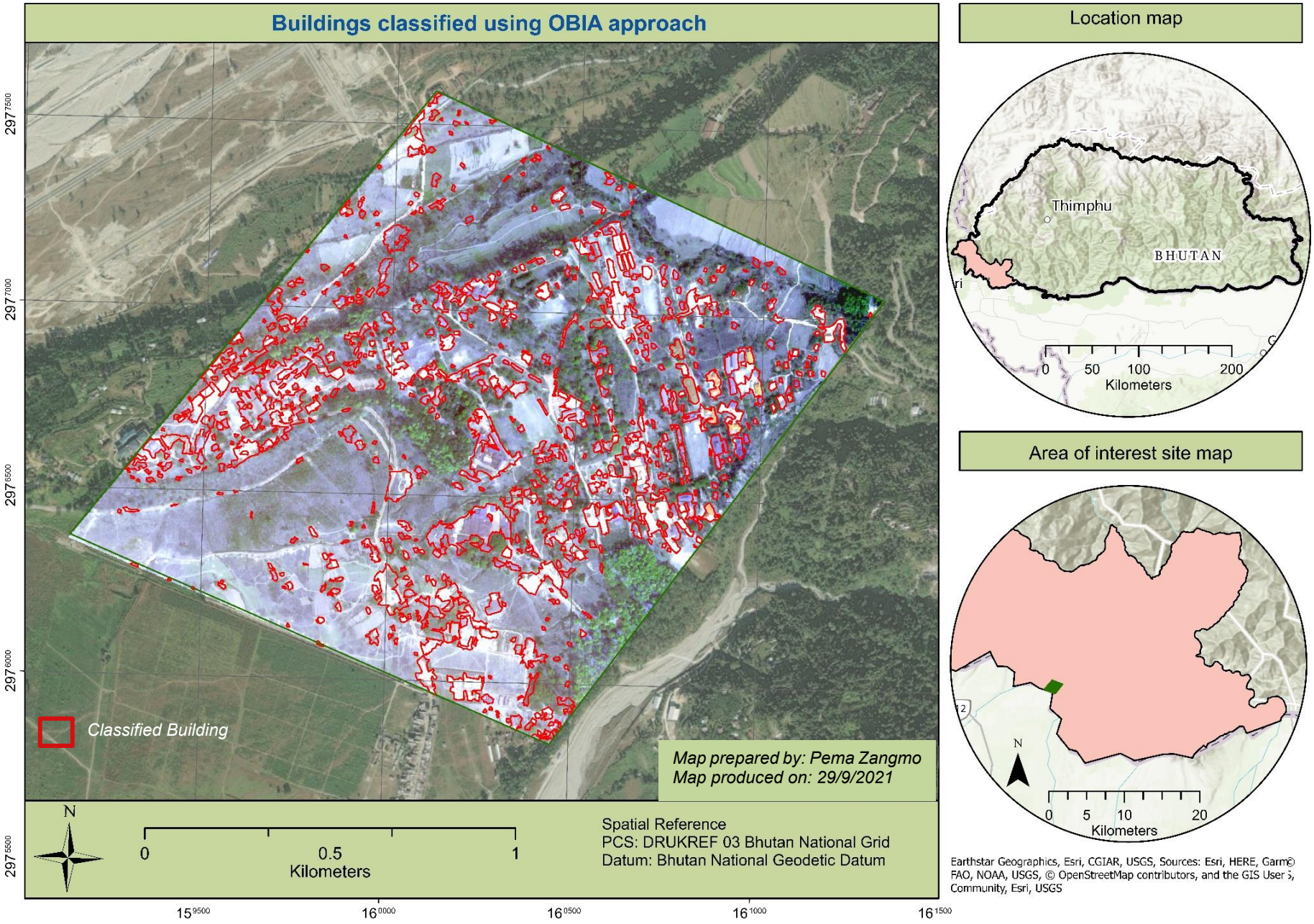
The way forward from such project would be not only limited to delineation of illegal settlement but also will pave path for carrying out future investigation. In Bhutan, there are land conflicts and issues related to construction of residential areas on wetland (legal plots) which are lawfully supposed to be maintained and conserved for only for cultivation purposes but violation of such laws has also been a concern in recent decades. As discussed in pervious sections, approximately only 3% of the of the land holdings are under residential use and rest are either for cultivation or for other land uses which could be the probable reason for such illegal construction. Land for human settlement is very scarce and constitution mandates 70% of the land in Bhutan to be conserved under forest areas. Such future studies can address the identification of encroachments which fall within the legal parcels and can be validated by cadastral plot database which contains the details regarding the land use of the legal plots.

7 ANNEXURE MAPS

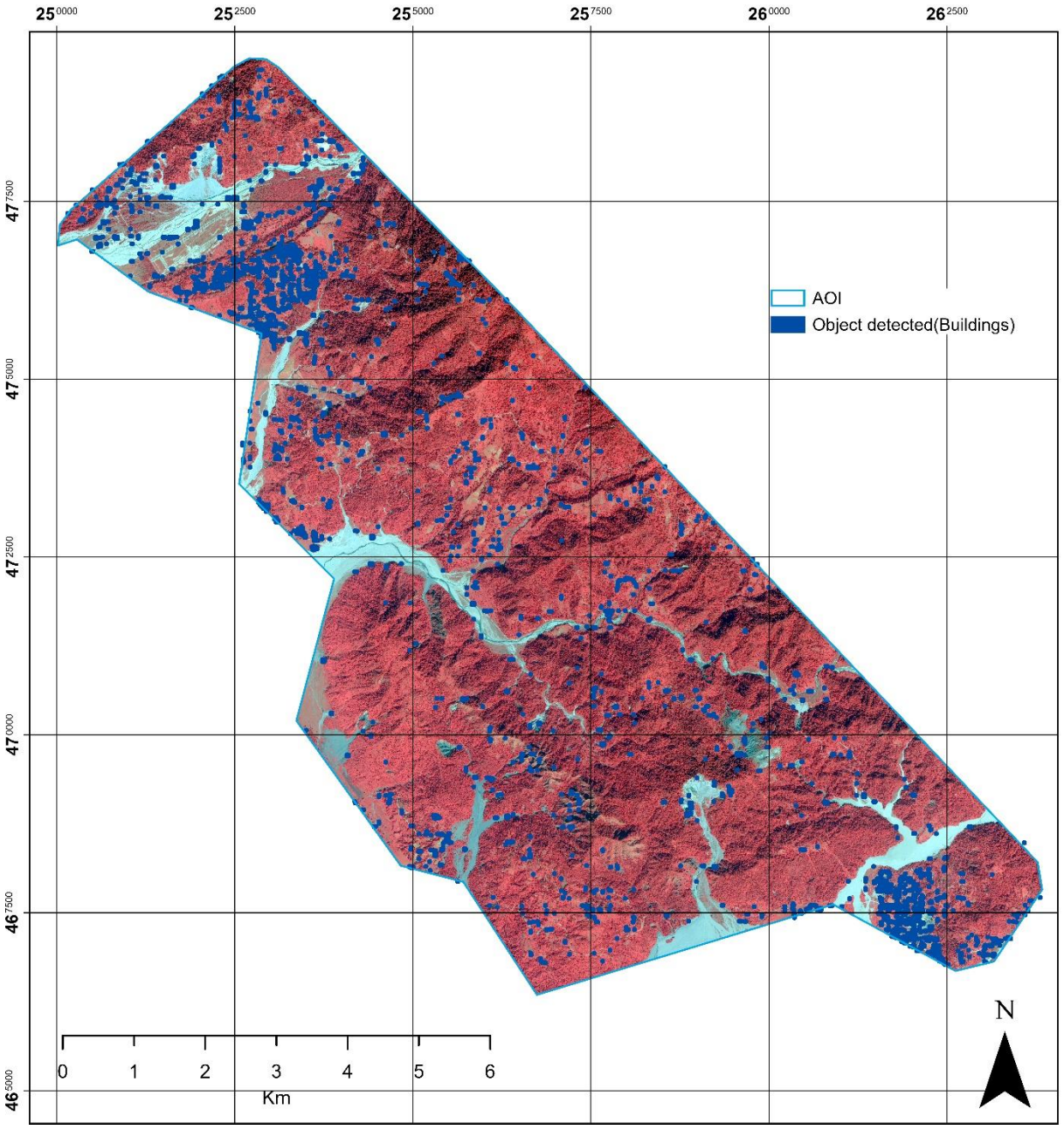
Annexure A. Segmented image using OBIA method



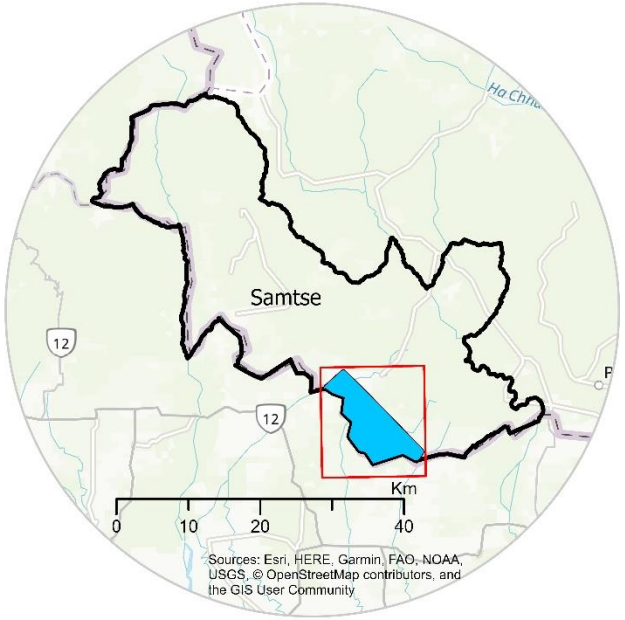
Annexure B. Building classified using OBIA Image classification



Annexure C.Building detected using Deep learning model -SSD model for Pleadies -1A imagery



Location Map



Spatial Reference
PCS: DRUKREF 03 Samtse TM
Datum: Bhutan National Geodetic Datum

Map prepared by : Pema Zangmo
FAN No: zang0038
Object detection carried out : 1/10/2021
Map produced : 10/10/2021

Object(buildings) Detection using Deep learning approach Single Shot Multi-Box Detector (SSD)-Pleaides-1A Imagery

8 REFERENCES

ACKER, J., WILLIAMS, R., CHIU, L., ARDANUY, P., MILLER, S., SCHUELER, C., VACHON, P. W. & MANORE, M. 2014. Remote Sensing from Satellites☆. *Reference Module in Earth Systems and Environmental Sciences.* Elsevier.

AGUILAR, M. A., AGUILAR, F. J., MAR SALDAÑA, M. D. & FERNÁNDEZ, I. 2012. Geopositioning accuracy assessment of GeoEye-1 panchromatic and multispectral imagery. *Photogrammetric Engineering & Remote Sensing,* 78**,** 247-257.

AGUILAR, M. A., DEL MAR SALDANA, M. & AGUILAR, F. J. 2013. Assessing geometric accuracy of the orthorectification process from GeoEye-1 and WorldView-2 panchromatic images. *International Journal of Applied Earth Observation and Geoinformation,* 21**,** 427-435.

AGUILAR, M. A., NEMMAOUI, A., AGUILAR, F. J., NOVELLI, A. & GARCÍA LORCA, A. 2017. Improving georeferencing accuracy of Very High Resolution satellite imagery using freely available ancillary data at global coverage. *International Journal of Digital Earth,* 10**,** 1055-1069.

ARROYO, C. M. 2011. Socio-spatial patterns: The backbone of informal settlement regeneration. *Proceedings of the Cities to be Tamed*.

ATTARZADEH, R. & MOMENI, M. 2012. Object-Based Building Extraction from High Resolution Satellite Imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XXXIX-B4.

AVIS, W. R. 2016. Urban Governance. Birmingham, UK: GSDRC, University of Birmingham.

BASAEED, E., BHASKAR, H. & AL-MUALLA, M. Comparative analysis of pan-sharpening techniques on DubaiSat-1 images. Proceedings of the 16th International Conference on Information Fusion, 9-12 July 2013 2013. 227-234.

BASRI, H., SYARIF, I. & SUKARIDHOTO, S. Faster R-CNN Implementation Method for Multi-Fruit Detection Using Tensorflow Platform. 2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC), 29-30 Oct. 2018 2018. 337-340.

BELGIU, M. & DRĂGUŢ, L. 2014. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. *ISPRS Journal of Photogrammetry and Remote Sensing,* 96**,** 67-75.

BENARCHID, O., RAISSOUNI, N., SAMIR, E. A., EL ABBOUS, A., AZYAT, A., BEN ACHHAB, N., LAHRAOUA, M. & ASAAD, C. 2013. Building Extraction using Object-Based Classification and shadow Information in Very High Resolution Multispectral Images, a Case Study: Tetuan, Morocco. *Canadian Journal on Image Processing and Computer Vision,* 4**,** 1-8.

BHUTAN, N. S. B. O. 2017. Bhutan living standards survey report. *In:* NATIONAL STATISTICS BUREAU, R. G. O. B. (ed.). Bhutan

BLASCHKE, T. 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing,* 65**,** 2-16.

BLASCHKE, T., LANG, S. & HAY, G. 2008. *Object-based image analysis: spatial concepts for knowledge-driven remote sensing applications*, Springer Science & Business Media.

BUREAU, B. N. S. 2020. Statistical Yearbook of Bhutan 2020. Thimphu,Bhutan.

CAO, M.-T., TRAN, Q.-V., NGUYEN, N.-M. & CHANG, K.-T. 2020. Survey on performance of deep learning models for detecting road damages using multiple dashcam image resources. *Advanced Engineering Informatics,* 46**,** 101182.

CHMIEL, J., KAY, S. & SPRUYT, P. ORTHORECTIFICATION AND GEOMETRIC QUALITY ASSESSMENT OF VERY HIGH SPATIAL RESOLUTION SATELLITE IMAGERY FOR COMMON AGRICULTURAL POLICY PURPOSES. 2004.

DARE, P. M. & FRASER, C. S. 2010. Cover: Mapping informal settlements using high resolution satellite imagery. *International Journal of Remote Sensing,* 22**,** 1399-1401.

DDA , D. D. A. 2007. Master Plan for Delhi-2021. Delhi,India.

DOVEY, K., SHAFIQUE, T., VAN OOSTRUM, M. & CHATTERJEE, I. 2020. Informal settlement is not a euphemism for 'slum': what's at stake beyond the language? *International Development Planning Review***,** 1-12.

DUBOVYK, O., SLIUZAS, R. & FLACKE, J. 2011. Spatio-temporal modelling of informal settlement development in Sancaktepe district, Istanbul, Turkey. *ISPRS Journal of Photogrammetry and Remote Sensing,* 66**,** 235-246.

ESRI. *Accuracy Assessment* [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/help/analysis/image-analyst/accuracy-assessment.htm [Accessed 2/10 2021].

ESRI. *Fundamentals of georeferencing a raster dataset* [Online]. Available: https://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/fundamentals-for-georeferencing-a-raster-dataset.htm [Accessed 29/9 2021].

ESRI. *Overview of image classification* [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/help/analysis/image-analyst/overview-of-image-classification.htm [Accessed 1/10 2021].

ESRI. 2021. *How the Compute Accuracy For Object Detection tool works* [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/how-compute-accuracy-for-object-detection-works.htm [Accessed 13/10 2021].

FALLATAH, A. M. O. 2020. *Mapping Informal Settlements in a Middle Eastern Environment Using Remote Sensing Techniques.* RMIT University.

GAVANKAR, N. L. & GHOSH, S. K. 2018. Automatic building footprint extraction from high-resolution satellite image using mathematical morphology. *European Journal of Remote Sensing,* 51**,** 182-193.

GEVAERT, C. M., PERSELLO, C., SLIUZAS, R. & VOSSELMAN, G. 2017. Informal settlement classification using point-cloud and image-based features from UAV data. *ISPRS Journal of Photogrammetry and Remote Sensing,* 125**,** 225-236.

GRAESSER, J., CHERIYADAT, A., VATSAVAI, R. R., CHANDOLA, V., LONG, J. & BRIGHT, E. 2012. Image Based Characterization of Formal and Informal Neighborhoods in an Urban Landscape. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 5**,** 1164-1176.

GRAM-HANSEN, B. J., HELBER, P., VARATHARAJAN, I., AZAM, F., COCA-CASTRO, A., KOPACKOVA, V. & BILINSKI, P. 2019. Mapping Informal Settlements in Developing Countries using Machine Learning and Low Resolution Multi-spectral Data. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society.*

HACKELOEER, A., KLASING, K., KRISP, J. M. & MENG, L. 2014. Georeferencing: a review of methods and applications. *Annals of GIS,* 20**,** 61-69.

HAGEN, E. 2011. Mapping Change: Community Information Empowerment in Kibera (Innovations Case Narrative: Map Kibera). *Innovations: Technology, Governance, Globalization,* 6**,** 69-94.

HASAN, A. 2016. Orangi Pilot Project: the expansion of work beyond Orangi and the mapping of informal settlements and infrastructure. *Environment and Urbanization,* 18**,** 451-480.

HNATUSHENKO, V., MOZGOVOY, D. & VASYLIEV, V. 2018. *Accuracy evaluation of automated object recognition using multispectral aerial images and neural network*.

HOESER, T., BACHOFER, F. & KUENZER, C. 2020. Object detection and image segmentation with deep learning on Earth observation data: A review—Part II: Applications. *Remote Sensing,* 12**,** 3053.

HOESER, T. & KUENZER, C. 2020. Object detection and image segmentation with deep learning on earth observation data: A review-part i: Evolution and recent trends. *Remote Sensing,* 12**,** 1667.

III, U. N. T. T. O. H. 2015. HABITAT III Issue papers 22 – informal settlements. New York.

JACOBSEN, K. 2011. Characteristics of very high resolution optical satellites for Topographic mapping. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XXXVIII-4/W19.

JAIN, S. 2007. Use of IKONOS satellite data to identify informal settlements in Dehradun, India. *International Journal of Remote Sensing,* 28**,** 3227-3233.

JAWAK, S. & LUIS, A. 2013. A Comprehensive Evaluation of PAN-Sharpening Algorithms Coupled with Resampling Methods for Image Synthesis of Very High Resolution Remotely Sensed Satellite Data. *Advances in Remote Sensing,* 2**,** 32-344.

KAVZOGLU, T., BILÜCAN, F. & TEKE, A. 2020. *COMPARISON OF SUPPORT VECTOR MACHINES, RANDOM FOREST AND DECISION TREE METHODS FOR CLASSIFICATION OF SENTINEL - 2A IMAGE USING DIFFERENT BAND COMBINATIONS*.

LEE, C., KIM, H. J. & OH, K. W. Comparison of faster R-CNN models for object detection. 2016 16th International Conference on Control, Automation and Systems (ICCAS), 16-19 Oct. 2016 2016. 107-110.

LI, J., LI, Y., CHAPMAN, M. A. & RÜTHER, H. 2005. Small format digital imaging for informal settlement mapping. *Photogrammetric Engineering & Remote Sensing,* 71**,** 435-442.

LIU, W., ANGUELOV, D., ERHAN, D., SZEGEDY, C., REED, S., FU, C.-Y. & BERG, A. C. SSD: Single Shot MultiBox Detector. *In:* LEIBE, B., MATAS, J., SEBE, N. & WELLING, M.,

eds. Computer Vision – ECCV 2016, 2016// 2016 Cham. Springer International Publishing, 21-37.

MADHAVAN, E. 2017. Building Identification in Satellite Images Using ANFIS Classifier. *Asian Journal of Applied Science and Technology (AJAST) Volume,* 1**,** 283-285.

MARANGOZ, A., ORUC, M., KARAKıŞ, S. & SAHIN, H. 2006. Comparison of Pixel-Based and Object-Oriented Classification Using Ikonos Imagery for Automatic Building Extraction–Safranbolu Testfield.

MARCELLO, J., MEDINA, A. & EUGENIO, F. 2013. Evaluation of Spatial and Spectral Effectiveness of Pixel-Level Fusion Techniques. *IEEE Geoscience and Remote Sensing Letters,* 10**,** 432-436.

MAXWELL, A. E., WARNER, T. A. & GUILLÉN, L. A. 2021a. Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review. *Remote Sensing,* 13.

MAXWELL, A. E., WARNER, T. A. & GUILLÉN, L. A. 2021b. Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 2: Recommendations and Best Practices. *Remote Sensing,* 13.

MEINEL, G. & NEUBERT, M. 2004. A comparison of segmentation programs for high resolution remote sensing data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* 35.

MOWHS 2016. The 3rd UN Conference on Housing and Sustainable Urban Development. *In:* SETTLEMENT, M. O. W. A. H. (ed.). Bhutan Royal Government of Bhutan.

NIEMEYER, J., ROTTENSTEINER, F. & SOERGEL, U. 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS journal of photogrammetry and remote sensing,* 87**,** 152-165.

NLC, N. L. C. 2012. Empowering land governance through sustainable Geo-Information management.

NORMAN, M., MOHD SHAHAR, H., MOHAMAD, Z., RAHIM, A., AMRI MOHD, F. & ZULHAIDI MOHD SHAFRI, H. 2021. Urban building detection using object-based image analysis (OBIA) and machine learning (ML) algorithms. *IOP Conference Series: Earth and Environmental Science,* 620**,** 012010.

NUSSBAUM, S. & MENZ, G. 2008. *Object-Based Image Analysis and Treaty Verification: New Approaches in Remote Sensing – Applied to Nuclear Facilities in Iran*.

ODONGO, M. A. 2017. *Use of open source Geo-solutions to develop a Cadastral model for informal settlement in Nairobi. Case study of redeemed village in Huruma informal settlement.* University of Nairobi.

OJWANG, I. O. 2009. *Spatial analysis of informal settlement sprawl and its environmental impact: a case study of Kibera.* University of Nairobi.

OKONGO, T. M. 2019. *Development of a Web-based Informal Cadastre for an Informal Settlement in Nairobi A Case Study of Mukuru Sinai Settlement.* University of Nairobi.

PADILLA, R., NETTO, S. L. & SILVA, E. A. B. D. A Survey on Performance Metrics for Object-Detection Algorithms. 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), 1-3 July 2020 2020. 237-242.

PARK, I. & KIM, S. Performance Indicator Survey for Object Detection. 2020 20th International Conference on Control, Automation and Systems (ICCAS), 13-16 Oct. 2020 2020. 284-288.

PUSHPARAJ, J. & HEGDE, A. V. 2017. Evaluation of pan-sharpening methods for spatial and spectral quality. *Applied geomatics,* 9**,** 1-12.

SALMAN, S. & LIU, X. 2019. Overfitting mechanism and avoidance in deep neural networks. *arXiv preprint arXiv:1901.06566*.

SAMPER, J., SHELBY, J. A. & BEHARY, D. 2020. The Paradox of Informal Settlements Revealed in an ATLAS of Informality: Findings from Mapping Growth in the Most Common Yet Unmapped Forms of Urbanization. *Sustainability,* 12.

SHACKELFORD, A. K., DAVIS, C. H. & WANG, X. Automated 2-D building footprint extraction from high-resolution satellite multispectral imagery. IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, 2004. IEEE, 1996-1999.

SIETCHIPING, R. Prospective slum policies: conceptualization and implementation of a proposed informal settlement growth model. Third Urban Research Symposium.“Land, Urban and Poverty Reduction, 2005. Citeseer, 4-6.

SPACE, A. D. A. 2021. Pléiades Imagery. Airbus Defence and Space.

SUN, W., CHEN, B. & MESSINGER, D. 2014. Nearest-neighbor diffusion-based pan-sharpening algorithm for spectral images. *Optical Engineering,* 53**,** 013107.

THOMSON, D. R., KUFFER, M., BOO, G., HATI, B., GRIPPA, T., ELSEY, H., LINARD, C., MAHABIR, R., KYOBUTUNGI, C., MAVITI, J., MWANIKI, D., NDUGWA, R., MAKAU, J., SLIUZAS, R., CHERUIYOT, S., NYAMBUGA, K., MBOGA, N., KIMANI, N. W., DE ALBUQUERQUE, J. P. & KABARIA, C. 2020. Need for an Integrated Deprived Area "Slum" Mapping System (IDEAMAPS) in Low- and Middle-Income Countries (LMICs). *Social Sciences,* 9.

TOUTIN, T. 2004. Review article: Geometric processing of remote sensing images: models, algorithms and methods. *International Journal of Remote Sensing,* 25**,** 1893-1924.

TOUZANI, S., PRITONI, M., SINGH, R. & GRANDERSON, J. 2020. Machine Learning for Automated Extraction of Building Geometry. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).

UN-HABITAT 2007. Sustainable Urbanization: local action for urban poverty reduction, emphasis on finance and planning. Nairobi, Kenya.

UN-HABITAT, U. N. H. S. P. 2020. World Cities Report 2020- The Value of Sustainable Urbanization.

VAIDYA, B. & PAUNWALA, C. 2019. Deep learning architectures for object detection and classification. *Smart Techniques for a Smarter Planet.* Springer.

VEERANAMPALAYAM SIVAKUMAR, A. N., LI, J., SCOTT, S., PSOTA, E., J. JHALA, A., LUCK, J. D. & SHI, Y. 2020. Comparison of Object Detection and Patch-Based Classification Deep Learning Models on Mid- to Late-Season Weed Detection in UAV Imagery. *Remote Sensing,* 12.

XIN, H. & LIANGPEI, Z. 2012. Morphological Building/Shadow Index for Building Extraction From High-Resolution Imagery Over Urban Areas. *IEEE journal of selected topics in applied earth observations and remote sensing,* 5**,** 161-172.

YU, L. & GONG, P. 2012. Google Earth as a virtual globe tool for Earth science applications at the global scale: progress and perspectives. *International Journal of Remote Sensing,* 33**,** 3966-3986.

YUSUF, Y., ALIMUDDIN, I., SRI SUMANTYO, J. & KUZE, H. 2012. Assessment of pan-sharpening methods applied to image fusion of remotely sensed multi-band data. *International Journal of Applied Earth Observation and Geoinformation,* 18**,** 165–175.

ZHANG, D.-D., ZHANG, L., ZABOROVSKY, V., XIE, F., WU, Y.-W. & LU, T.-T. 2019. *Research on the pixel-based and object-oriented methods of urban feature extraction with GF-2 remote-sensing images*.

ZHANG, Y. & MISHRA, R. K. A review and comparison of commercially available pan-sharpening techniques for high resolution satellite image fusion. 2012 IEEE International geoscience and remote sensing symposium, 2012. IEEE, 182-185.

ZHAO, K., KANG, J., JUNG, J. & SOHN, G. Building extraction from satellite images using mask R-CNN with building boundary regularization. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018. 247-251.