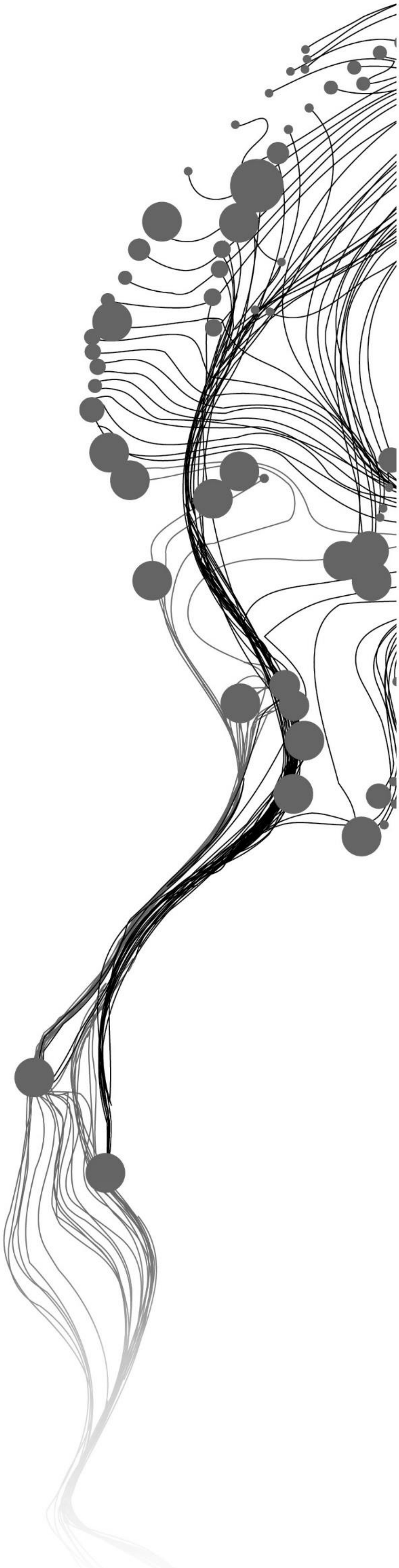
ANALYZING URBAN DEVELOPMENT PATTERNS IN A CONFLICT ZONE: A CASE STUDY OF KABUL

VINEET CHATURVEDI
July, 2019

SUPERVISORS:
dr. M. Kuffer (Monika)
dr. D. Kohli (Divyani)



ANALYZING URBAN DEVELOPMENT PATTERNS IN A CONFLICT ZONE: A CASE STUDY OF KABUL

VINEET CHATURVEDI
Enschede, The Netherlands, July 2019

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Urban Planning and Management

SUPERVISORS:
dr. M. Kuffer (Monika)
dr. D. Kohli (Divyani)

THESIS ASSESSMENT BOARD:
Prof. Dr. P.Y. Georgiadou (Chair)
dr. Maik Netzband (External Examiner, University of Würzburg)

ABSTRACT

To measure the development of a city in conflict zone is a daunting task because of the lack of access to the field data and security issues. This research aims to analyse the patterns of growth and the factors influencing the growth in the city of Kabul with the help of high-resolution earth observation-based data using logistic regression model. The growth is analysed from 2001 to 2017. The research explores the possibility to analyse settlement patterns by extracting image features from high resolution aerial photograph and terrain features as input to evaluate the capabilities of Random Forest classifier. The complexity of urban morphology in the city of Kabul made it difficult to classify the different settlement patterns due to the similarities in the texture of the planned and unplanned settlements and low contrast between the built and non-built. The texture similarity composed of localities consisting of mud houses and barren land also made it difficult to separate built-up from other classes. The classification consisted of 7 classes which includes 4 categories of built-up class, water, vegetation and barren land. The built-up is classified on the basis of built-up densities into medium density planned, low density planned, hillside settlements unplanned and medium density unplanned. The overall accuracy achieved is 50% with lowest accuracies for classes medium density planned and low density planned. To improve the accuracy of the classification the 2 built-up classes for the planned settlements are merged and the 2 built-up classes for unplanned are merged and the overall accuracy achieved is 61%. The unplanned settlements have grown by nearly 4.5 times from 2001 to 2017 whereas planned settlements have grown by 1.25 times. The growth in unplanned settlements is mostly towards the west and north west parts of the city and the growth in planned settlements is in the central and eastern parts of the city. The research further analyses the factors that influence the growth patterns in a conflict zone using binary logistic regression model. To perform the regression the change in the planned and unplanned settlements in Kabul is measured using the classified aerial photograph of 2017 and the Built-up layer extracted from IKONOS for the year 2001. The results of the regression reveal that population density and military bases are important factor for both planned and unplanned settlements. Slope is also one of the factors for the growth of unplanned settlements, which means there is growth of unplanned settlements on steeper slopes on the hillside.

ACKNOWLEDGEMENTS

I take this opportunity to thank my supervisors dr. Monika Kuffer and dr. Divyani Kohli for encouraging me to take up this topic for my MSc research and to provide consistent guidance and support throughout my research. It would have been impossible to carry out the research without their valuable inputs. It has been a great learning experience and a good opportunity to expand my knowledge in machine learning, image processing and urban growth modelling.

I would like to thank ITC excellence scholarship for funding my studies and making my stay at ITC, Enschede a memorable one. A big thank you to all professors, students, staff of the department of Urban Planning and Management. My special thanks to my friends Mitali, Logambal, Xia Xue, Maryam, Tanvir, and Maliwa who have been quite helpful and supportive throughout my stay.

My special thanks to JRC and Farhad MiaKhel from CRIDA (Capital Regions Independent Development Authority) for providing me the required data to carry out my research.

Lastly, I would also like to thank my father Anil Mohan and my sister Puneet for keeping my spirits high at all times.

TABLE OF CONTENTS

1. Introduction 1
   1.1. Background and Justification 1
   1.2. Research problem 2
   1.2.1. Conceptual Framework 3
   1.2.2. Thesis structure 4
2. Literature review 5
   2.1. Studies on modeliing of urban growth dynamics in Afghanistan 5
   2.2. Machine learning and Urban growth modelling 5
   2.3. Modeling urban growth dynamics and land use change in other parts of the world 6
3. Study Area, data and methodology 9
   3.1. Kabul faces the following urban development challenges 10
   3.2. Data 11
   3.3. Methodology 12
   3.4. Classification 13
   3.4.1. Random Forest Classification 15
   3.5. Spatial Logistic Regression 18
4. Results and Discussion 21
   4.1. Classification Results 21
   4.1.1. Feature Selection 21
   4.2. Accuracy Assessment 25
   4.3. Results of the Logistic Regression 27
   4.4. Amount and patterns of growth in Planned and Unplanned Settlements 27
   4.5. Binary Logistic Regression 28
   4.6. Discussion 39
5. Conclusion And recommendation 41
6. References 43

LIST OF FIGURES

Figure 1 Afghanistan GDP annual growth rate 2
Figure 2 Conceptual Frame 3
Figure 3 Kabul city map with 22 districts and Land use classes, Source: Atlas of Afghan city regions, 20169
Figure 4 Flowchart showing the methodology 12
Figure 5 Attacked locations 20
Figure 6 Training samples 22
Figure 7 Reference samples 22
Figure 8 Permutation Importance 23
Figure 9 GINI Decrease 23
Figure 10 Result of Random Forest Classification with 7 classes 24
Figure 11 Results of the classification with 5 classes 24
Figure 12 Changed in planned and unplanned settlement from 2001 to 2017 27
Figure 13 Predictor variables for regression 30
Figure 14 ROC curve for Binary logistic regression on change in planned settlement 37
Figure 15 ROC curve for Binary logistic regression on change in unplanned settlement 38
Figure 16 Global Human Settlement Layer 39

LIST OF TABLES

Table 1 Data collected and its sources 11
Table 2 Classification of Built-up and Non-Built-Up 13
Table 3List of image features included in the Random Forest Classification 16
Table 4 List of spectral Indices 17
Table 5 List of factors influencing patterns of urban growth 18
Table 6 Summary of the classification accuracy with seven classes 25
Table 7 User and producer accuracy of Individual classes 25
Table 8 Summary of Classification Accuracy 25
Table 9 User and producer accuracy for individual classes 25
Table 10 Accuracy results of classification with Built and Vegetation, barren and Water class 26
Table 11 Accuracy result of the combined built-up and Water class 26
Table 12 Comparison of the overall accuracy of Global Human Settlement layer and Classification 26
Table 13: Growth in the planned and unplanned settlements 28
Table 14 List of Variables for the logistic regression 28
Table 15 Pearson's coefficient correlation for variables for change in planned settlements 31
Table 16 Variance Inflation Factor for the predictor variables for change in planned settlements 32
Table 17 VIF after removing variables with very high values for change in planned settlements 32
Table 18 Significance, Odds Ratio, Wald statistics results of the regression for planned settlements 33
Table 19 Model summary for regression on change in planned settlements 33
Table 20 Classification table results for change in planned settlements 33
Table 21 Pearson's coefficient correlation for variables for change in unplanned settlements 34
Table 22 Variance Inflation Factor for the predictor variables for change in unplanned settlements 35
Table 23 VIF after removing variables with very high values for change in unplanned settlements 35
Table 24 Significance, Odds Ratio, Wald statistics results of the regression for unplanned settlements ... 36
Table 25 Model summary for regression on change in unplanned settlements 36
Table 26 Classification table results for change in unplanned settlements 36
Table 27 Area under ROC curve for regression on change in planned settlements 37
Table 28 Area under ROC curve for regression on change in unplanned settlement 38
Table 29 Factors with the highest odds ratio for change in the planned settlements 40

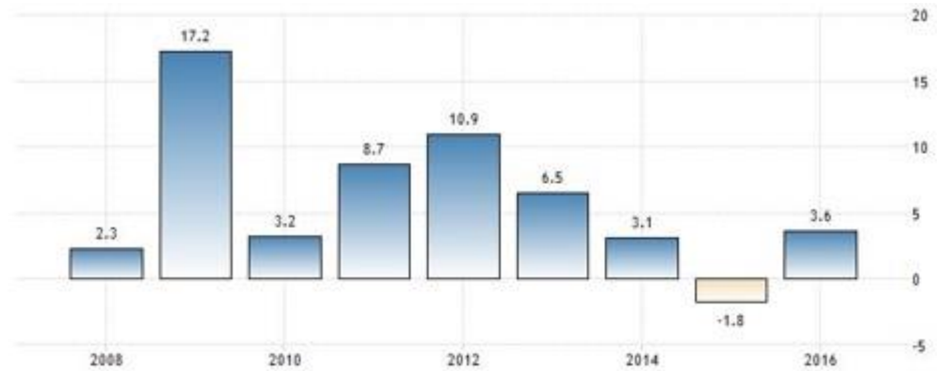
1. INTRODUCTION

The fast pace of urbanization has resulted in the change of land use and land cover which has affected the physical environment. Urban growth is a complex process which involves various geophysical, social, environmental, planning, government policies and economic factors etc. In order to understand the dynamics of urban growth it is necessary to assess the urban development trend and what are the causes of this development. Historically many models have been applied to study urban dynamics. Some of these models include Von Thunen's land use model and urban growth theory which explains how the market forces control the spatial distribution of land use and urban growth, in 1926 Burgess proposed the concentric zone theory, in 1933 Walter Christaller proposed the central place theory, the sector theory, the multiple nuclei theory and finally the bid rent theory (Pradhan, 2017). However, the dynamics and continuous growth of cities has increased the complexity of modelling urban growth. In recent years advanced modelling techniques have been developed using Geographical information system (GIS), Remote Sensing and statistical techniques to understand and explain the urban environment. These models have been applied on many cities to understand patterns of urban growth and the factors influencing this growth however there are very few studies carried out in understanding this phenomenon in the context of a conflict zone. The present research analyses the patterns of urban growth in a conflict zone using multi-temporal data and machine learning to map the growth and analyse the important factors that drives the urban growth and how does specific factors relate to being in a conflict zone. The availability of high-resolution satellite images and aerial photographs allows to carry out spatio-temporal analysis of urban growth in a data poor environment with challenges related to security and safety. Random Forest (RF) classification has been employed using high resolution aerial photographs and Spatial Logistic Regression (SLR) modelling to analyse the patterns of urban growth in the city of Kabul. SLR consists of statistical analysis methods such as multiple regression and logistic regression and are widely used in urban growth and pattern modelling. These methods have been successfully applied to find out the important predictor variables impacting urban growth and to understand the spatial phenomena of urban development (Cheng & Masser, 2003).

1.1. Background and Justification

In 1950, only 30 per cent of the world's population lived in urban areas, a proportion that grew to 55 per cent by 2018 (United Nations, 2018). Some of the reasons for this increase in population is tremendous amount of infrastructure investment, availability of job opportunities, better living standards, relatively secure environment. However, the pace of growth is not the same everywhere. War zones like Afghanistan are facing tremendous challenges involving the reconstruction of their cities. There are many factors that influence the growth of cities in a conflict zone. Afghanistan is one such country that has experienced unprecedented growth in urban population. After the US invasion in October 2001 when the US and allied forces drove the Taliban out of power, all the major cities in Afghanistan have experienced significant population growth (*Migration and Urban Development in Kabul: Classification or Accommodation*, 2012). Most cities are experiencing influx of returned refugees and internally displaced persons. Afghanistan is also one of the poorest countries in the world in terms of human development (Asian Development Bank, 2017). This rapid urbanization combined with a very limited capacity of local authorities to deal with this growth leads to increase in poor and vulnerable population, which poses a challenge in urban planning and management. Post war aid and humanitarian assistance in Afghanistan as part of the state building efforts has not only had an impact on the Gross Domestic Product (GDP) but also on the direction of growth of some of its major cities. According to Bizhan (2018), the GDP in Afghanistan grew annually by 10% on average between 2002 and 2013. The Gross Domestic Product (GDP) expanded 3.60 percent in 2016 from the previous year. GDP Annual Growth Rate averaged 6.98 percent from 2007 until 2016, reaching a record high of 17.20 percent in 2009 and a record low of -1.80 percent in 2015 ("Afghanistan GDP Annual Growth Rate," n.d.). as shown in Fig 1.

Central Statistics Organisation of Afghanistan: GDP annual growth rate



Source: tradingeconomics.com

Figure 1 Afghanistan GDP annual growth rate

In the context of the rapid urbanization in Afghanistan, the growth of cities without proper planning and management has resulted in fragmented development (Jochem, Bird, & Tatem, 2018). Owing to the large number of donor funded development projects, investing especially into infrastructure like roads, energy, sanitation and other basic services like schools, hospitals, commercial areas etc., the pace of urbanization has accelerated. Different donor funded agencies and organizations like USAID, ADB, World Bank and the US Army corps of Engineers have invested in these infrastructure projects to improve the situation of services and amenities, "economic infrastructure and services include sectors such as energy (1.93 billion USD from 2011-2015), and transport and communications (1.23 billion USD)". "Over 2 billion USD has been spent in Afghanistan from 2011-2015"(*Aid Effectiveness in Afghanistan*, 2018). For example, "USAID has invested in the construction and rehabilitation of more than 2,000 kilometres (km) of roads to propel travel and commerce, including Afghanistan's Ring Road, which connects the country's five major cities: Herat, Kabul, Kandahar City, Jalalabad, and Mazar-e-Sharif. More than 80 percent of Afghans now live within 50 km of the Ring Road"("Infrastructure," 2018). "Investments were made in infrastructure in particularly conflict-ridden areas rather than in the people of Afghanistan" (Mielke, Bicc, & Grawert, 2016). The international donor funded investments in the cities has created a huge gap between the living standards of people living in the rural and urban areas. A large number of rural and internally displaced Afghan are flocking to the cities in search of jobs and better standard of living. This has led to a very large unplanned urban growth. The migration of refugees and internally displaced people has been putting immense pressure on the available resources in the city. Apart from economic growth major cities offer more safety than their original place of habitation.

1.2. Research problem

The post war city of Kabul has experienced massive population growth partially driven by the investments in several infrastructure development projects by various international organization as part of the post war reconstruction process, migration of return refugees, internally displaced people and safety and security. The existing and the newly developed infrastructure and services have not been able to serve the majority of the population of the city especially inhabitants living in the unplanned parts of the city. This has also resulted in not being able to provide adequate security for the whole population. Better infrastructure facilities will lead to a more secure environment. For example, the response time of all security measures such as ambulances, fire extinguishers, police, national security organizations will be faster during volatile situation if adequate infrastructure is in place. To provide all the necessary infrastructure for such a large population in an insecure and unstable situation is a challenge faced by the Government and other International agencies helping in the infrastructure development. The research is an attempt to analyse the important factors that influence the urban growth that are specific to conflict zone. Many studies exist that analyse the urban growth and drivers of urban growth but an analysis on the urban growth of a city in conflict has not been done. In the research "Evaluating Urban Land Expansion using Geographic Information System and Remote Sensing in Kabul city, Afghanistan" (Ahmadi & Kajita, 2016) the research mainly focuses on the relationship of factors like population growth, migration and economic growth with expansion of the urban land. Factors specific to a conflict zone have not been considered in the research. There are no research studies that have analysed the factors specific to conflict zones which influence the patterns of urban growth for the city of Kabul.

1.2.1. Conceptual Framework

To analyse spatial patterns of urban growth in a city ravaged by war and in a continuous state of insurgency requires the understanding of the factors that have influence this growth. This section deals with a conceptual frame which explains how a model can be developed which will explain the urban growth patterns in a war zone. According to the statistical data obtained from Central statistics office, Kabul the population of the city has increased from approximately 1.5 million in 2001 to around 4.9 million people by 2015 (Ahmadi & Kajita, 2017). Since 2001 the city has expanded, and new settlements have come up to accommodate the influx of immigrants. Most of the growth is unplanned within and in the outskirts of the city. All the hillside settlements are unplanned. There are some new townships coming up in the recent years built by private builders and developers. To capture the patterns of growth the scale of the data should be of high resolution. The temporal scale should have substantial gap to be able to analyse the change in the patterns of growth. The determinants of growth of the city should be specific to conflict zone. Figure 2 depicts the conceptual frame that would be followed to model the patterns of growth in the city of Kabul.

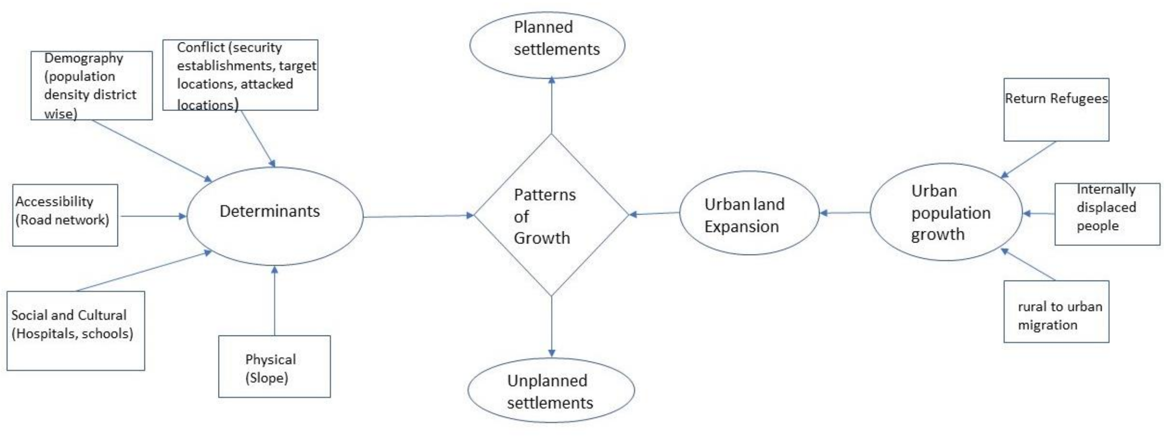


Figure 2 Conceptual Frame

Research Objectives and questions

The objective of the study is to analyse the patterns of urban growth in a conflict zone using multi temporal earth observation-based data.

- To map urban growth using multi temporal earth observation-based data
    - Which image features are most significant to map urban growth with machine learning?
    - What is the accuracy of the mono-temporal urban maps?
    - What has been the amount of urban growth in the last decade?
- To analyse physical, accessibility, conflict and socioeconomic factors of urban growth
    - What are the factors influencing urban growth specific to a conflict zone?
    - Which factors are most significant in explaining the urban growth patterns?
- To evaluate the model in explaining urban growth
    - What is the accuracy of the model?
    - How well do the drivers explain the observed urban growth patterns?
    - How successful is the model in explaining urban growth?

1.2.2. Thesis structure

The research is organized into 5 chapters

Chapter 1 is introduction to urban growth modelling, background and justification, research problem, conceptual framework, research objectives and questions

Chapter 2 is literature review which gives a brief on the various theories and modelling techniques applied in a conflict zone especially in Afghanistan. It also includes the various modelling approaches adopted to study urban dynamics in various cities around the world, the methodology used and their outcomes. It also highlights the shortcomings of these models.

Chapter 3 is Study area and Methodology this chapter introduces the city of Kabul and the growth of the city despite the ongoing conflict in the country. The chapter also covers the methodology adopted to carry out the processing and analysis of the data. It explains the Random forest classification process, selection of image feature and the selection of independent variables which are unique to a conflict zone for performing the logistic regression.

Chapter 4 is the **r**esults and discussion this chapter includes the results of classification of aerial photographs and accuracy assessment of the classification. It further analysis the results of the logistic regression model. The fitness of the model and how well it can explain the factors influencing patterns of urban growth. The discussion covers interpretation of the results and the significance of the outcomes. It also includes the shortcomings of the data used and what are the new findings achieved by using the logistic regression model.

Chapter 5 is **c**onclusion and recommendations. Conclusion describes the overall contribution of the study in achieving the objectives and the limitations of the study. Recommendation talks about future work and alternative methodologies and data sets that can be adopted for carrying out the research.

2. LITERATURE REVIEW

Many research studies have focussed on modelling urban growth patterns and the factors influencing this growth using different modelling methods. Relevant modelling methods, their results and limitations have been discussed in the chapter

2.1. Studies on modeliing of urban growth dynamics in Afghanistan

There are not many research studies carried out to study urban growth patterns of the cities in Afghanistan. Some studies have identified the settlement pattern in Kabul, mostly using satellite imagery and some studies using vector data. In the research "Identifying residential neighbourhood types from settlement points in a machine learning approach" geospatial vector data (i.e. point, line, polygon) is used to map land uses. This approach distinguished residential settlement types (regular vs irregular) using an existing database of settlement points locating structures. The method was tested in seven provinces of Afghanistan (Balkh, Helmand, Herat, Kabul, Kandahar, Kunduz, Nangarhar). Data features are calculated at multiple spatial scales. Land use classification is performed using the remote sensing measures of elevation, slope, vegetation and night-time lights using supervised machine learning method. The method is successful in identifying neighbourhood type however the computational cost is high. The classification was performed using both Random Forest and Maximum likelihood using the same training datasets. Random Forest model outperformed ML method even though ML method is simpler to implement, the RF method is able to find more complex and non-linear relationships which likely improves its performance (Jochem et al., 2018). There has been research done to "Evaluate Urban Land expansion using Geographic Information System and Remote Sensing in Kabul city, Afghanistan" (Ahmadi & Kajita, 2016). In the study they used temporal demographic data, temporal land use plan acquired from the master plan and satellite images from the 1964 to 2008 and to evaluate the temporal and spatial difference of urban expansion and land use categories. Urban land expansion is calculated using indicators like annual urban growth and elasticity of urban land expansion to urban population growth. The study analysis the relationship between growth of driving forces which are annual population growth rate, annual urban land growth rate and annual GDP growth rate affecting urban land expansion. Based on the data analysis the study found that the urban land has expanded 14 times from 1964 to 2008. The research paper "Evaluation of Urban Land Development Direction in Kabul City, Afghanistan" determines the direction of urban land development in Kabul using satellite data and vector data to analyse the urban growth and urban change (Ahmadi & Kajita, 2017). Research study is conducted for identification of formal and informal neighbourhoods in urban landscape using spatial, structural and contextual image feature. World-wide four cities were chosen for the research out of which Kabul and Kandahar were two cities from Afghanistan chosen as case studies (Graesser et al., 2012). A decision tree method was used to classify formal and informal settlements. A combination of low-level statistics is used in the study which can be computed at an acceptable speed. It was concluded that accuracy would have improved if statistics were computed with a moving window instead of discrete window.

2.2. Machine learning and Urban growth modelling

A major drawback of carrying out research on urban development in a conflict zone is lack of reliable data and the difficulties in carrying out field work for collection of data. Availability of aerial photographs / satellite imageries are the best source of data. Remotely sensed data are powerful tools for measuring and detecting several elements related to the urban morphology of metropolises, such as density, amount, textural form, shape, and diffusion of built-up areas(Pradhan, 2017). There are several machine learning algorithms which can be used to analyse urban growth patterns using remotely sensed data with which you can obtain higher accuracy results with limited field observations. Some of the algorithms used for image classification are Maximum Likelihood, Artificial Neural Network, Support Vector Machine, Decision tree and Random Forest. Machine learning approaches have been combining textural, spectral, and structural features, with studies confirming the remarkable performance of the random forest classifier (RFC) for extracting slums (Kuffer, Pfeffer, Sliuzas, & Baud, 2016). Object oriented classification methods are gaining popularity as compare to pixel-based classification as they are best suited for urban growth modelling studies. Regarding the occurrences of settlement areas in RS data, pixel-based approach on a high-resolution image cannot represent the heterogeneity of complex urban environment (Pradhan, 2017). Image feature measures like texture, spectral, contextual and structural have been utilized as inputs in various research for high resolution image classification to evaluate the performance of various machine learning algorithms. In the research paper “Exploring the Potential of Machine Learning for Automatic Slum Identification from VHR Imagery” (Duque, Patino, & Betancourt, 2017) various texture, spectral and structure feature were utilized for slum identification. Spectral features like NDVI have been used in research for classification. For land cover classification, some band combinations of the remote sensed data are exploited and the spatial distribution such as road, urban area, agriculture land and water resources are easily interpreted by computing their normalized difference vegetation index (Bhandari, Kumar, & Singh, 2012). In various research studies the Grey Level Co-occurrence Matrix (GLCM) has been employed for texture feature extraction as they are useful in pattern recognition.

The integration of GIS technology, Remote sensing data and secondary sources of data have been extensively utilized for urban growth and pattern modelling. Modelling of the complex spatial and temporal urban growth phenomena has gained popularity with the aid of faster computing power, availability of spatial data and improved algorithms. There are mainly two groups according to the key mechanisms to simulate the process of land use change rule-based/process-based models and empirical-statistic model (Nong & Du, 2011). Logistic regression is an empirical statistical model which studies transition of landcover from one state to another. Binary logistic regression can be used to predict a dependent variable on the basis of continuous and categorical independent variables and to determine the percent of variance in the dependent variable explained by the independent variables (Nong & Du, 2011).

2.3. Modeling urban growth dynamics and land use change in other parts of the world

Modelling of urban growth has been carried out for many cities around the world and there are several research papers focussed on land use change, urban patterns and factors influencing them. In the paper “Mapping patterns of urban development in Ouagadougou, Burkina Faso, using machine learning regression modelling with bi-seasonal Landsat time series” Schug et al., (2018) used Landsat-TM/ETM+/OLI time series data to quantity land area for the Ouagadougou metropolitan area between 2002 and 2013. In the research they used support vector regression with synthetically mixed training data. The research focusses on monitoring urban development on a neighbourhood scale by producing urban fractions and seasonal vegetation to soil ratio for mapping planned and unplanned development. The research was successful in adding a seasonal dimension to study of different types of urban patterns on the basis of identifying different landcover compositions.

The research paper “Modelling urban growth with GIS based cellular automata and least squares SVM rules: a case study in Qingpu–Songjiang area of Shanghai, China Yongjiu” (Feng, Liu, & Batty, 2016) presents machine learning CA model (termed Mach CA, meaning machine learning based CA) with non-linear transition rules based on least square support vector machine (LS-SVM) to simulate the spatial-temporal process of urban growth.

In the research paper "Urban growth pattern modelling: a case study of Wuhan city, PR China" (Cheng & Masser, 2003) the objective is to seek and compare determinants of urban growth pattern in this specific period. The major methodology developed here consists of exploratory and confirmatory data analysis. The statistical technique is a multivariate estimation method in examining the relative strength and significance of the factors (explanatory variables). The findings of the research are 1) urban growth obeys a law of negative exponential function in terms of probability of change 2) research also found that logistic regression analysis is very sensitive to its multi-stages such as data transformation and spatial sampling. 3)In this research, remotely sensed imagery is primary data sources for urban growth modelling although pattern analysis needs other socio- economic attributes 4) GIS is proven to be highly limited in spatial data analysis including exploratory and confirmatory.

Factors influencing the urban growth were studied in the paper "identifying factors influencing urban spatial growth for the George town conurbation" (Mahamud, Samat, & Noor, 2016). In the research data from recent research journals and articles regarding modelling urban growth were reviewed. In addition, an on-line survey was conducted to gather data on driving forces of urban development. The questionnaire consists of 5 items to gather information about respondents' demographic backgrounds and another 6 items to assess their knowledge of urban development. The survey contains open-ended and closed-ended questions using Likert scale rating 1-to-9. The paper concluded that political factor should be considered in predicting urban growth, but it is difficult to quantify it.

In the study "Urban Growth in the Bucharest Metropolitan Area: Spatial and Temporal Assessment Using Logistic Regression", Kucsicsa and Grigorescu (2018) used distance variables in explaining the urban growth. The authors used spatial data extracted from Landsat satellite images and map imagery software using geographical information system (GIS) techniques resulting in five thematic maps—built-up areas, roads, forests, water bodies, and major commercial centres. The study aims to explain the relationship between urban growth and the distance explanatory driving forces using logistic regression and GIS. The methodology used in the present study includes three main stages: (1) the extraction of the thematic maps from Landsat satellite images, (2) the spatial analysis using GIS, and (3) the statistical analysis using the Statistical Package for the Social Sciences (SPSS) software package. The binary logistic regression model indicates that urban growth has been driven mainly by the distance to nearest secondary roads, distance to existing built-up areas, distance to nearest major roads, distance to the nearest road junction, and distance to Bucharest. The findings of this approach confirm that variables related to distance have an important role in explaining urban growth.

Logistic regression has been applied in several studies to study the relationship between urban growth pattern and the driving factors influencing urban growth. One of the research studies "Analysing the spatial urban growth pattern by using logistic regression in Didim district" (Atak, Erdogan, Ersoy, & Nurlu, 2014) applied logistic regression model to model the determinants of urban patterns and to identify the relationship between urban growth, socio economic drivers and biophysical factors. The aim of the study is to generate a probability of urban growth map to analyse the urban growth patterns. Different scenarios of future urbanization patterns are predicted by allocation cells that represent the estimated size of the required urban land on the predicted probability map. Another research paper that demonstrate the utility of logistic regression in predicting urban growth is "Urban growth pattern modelling using logistic regression"(Nong & Du, 2011) it determines relationship between urban growth and driving factors and predicting where growth will occur. It is concluded in the paper that spatial autocorrelation has influence on the results carried out in logistic regression which needs further research.

.

3. STUDY AREA, DATA AND METHODOLOGY

The city of Kabul, capital and largest city of Afghanistan is located in the eastern part of the country at 33°56'20.8" N and 67°42'35.8" E and at an elevation of 1798 meters above sea level. Kabul is also a municipality forming a part of the Kabul Province. Kabul city is divided into 22 districts called "Nahiyas". A map of the city is shown in Fig 3.

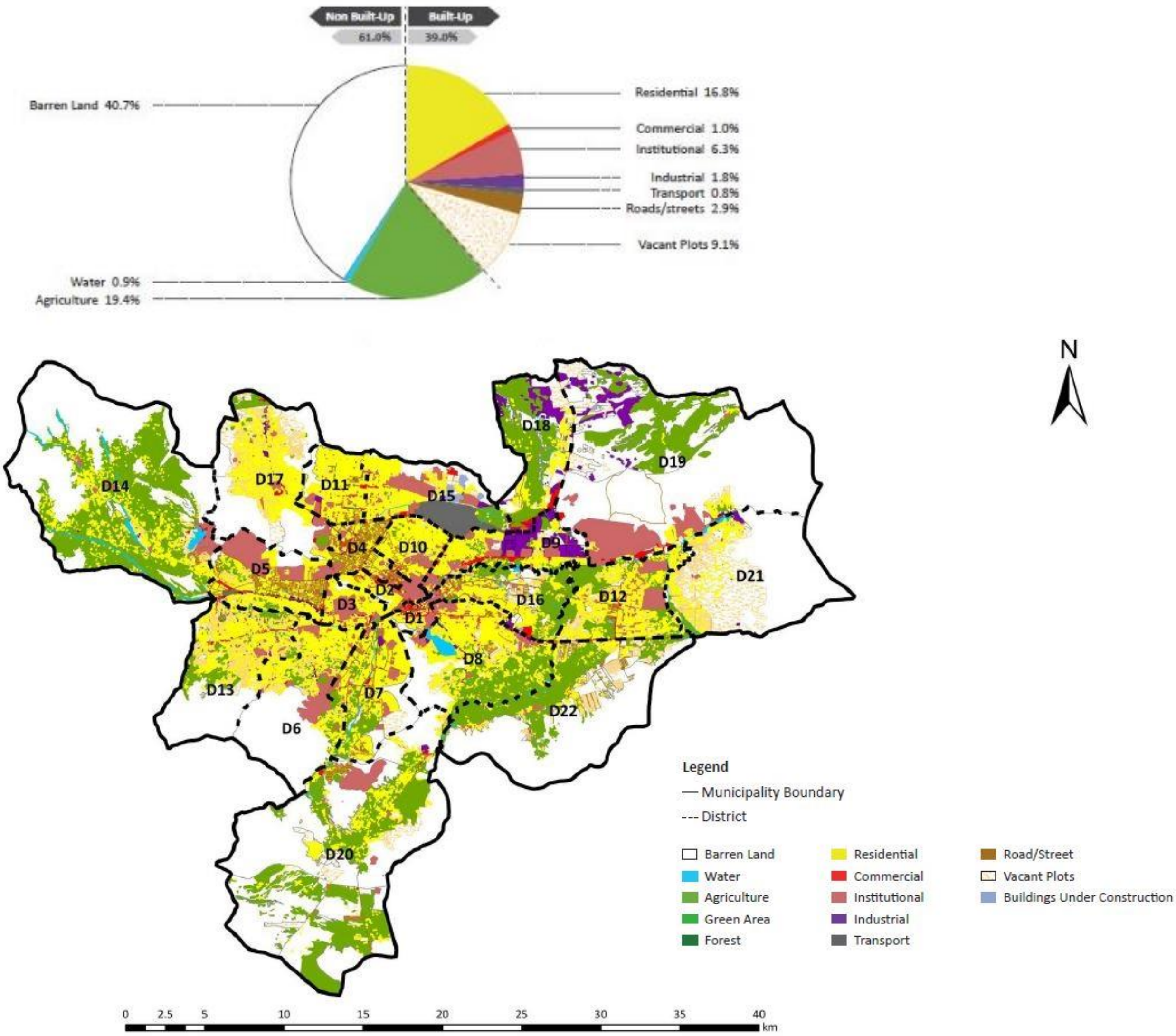


Figure 3 Kabul city map with 22 districts and Land use classes, Source: Atlas of Afghan city regions, 2016

Kabul is the fifth fastest growing city in the world, its population has grown from about 1.5 million in 2001 to around 5 million people in 2015. The rapid urbanization has not been able to keep up with demands of the growing population in a city which was originally designed for around 700,000 people. "An estimated 70% of Kabul's residents live in informal or illegal settlements" (Rasmussen, 2014). The total built-up area of Kabul city region is 40,143 ha out of which land area for residential use is 17,335 ha and for commercial, 1006 ha. (*Atlas of Afghan City Regions*, 2016). The primary factor and the driving force behind the urban land expansion of the city is economic growth (Ahmadi & Kajita, 2016). According to

World Bank (2018) the security situation in Afghanistan has been deteriorating following the withdrawal of international security forces in 2014 which has impacted the economic growth of the county. The safety dimension has led to infrastructure investment being limited to parts of the city leaving majority of the city unplanned which has given rise to large informal settlements on the hill sides and steep slopes.

3.1. Kabul faces the following urban development challenges

The city of Kabul has been a witness to a continuous conflict situation. The city has experienced tremendous expansion in last 2 decades which has gone beyond its municipal limits. Kabul's population has been growing and will continue to grow in the coming years. The city is facing various urban issues related to sustenance of its growing population. Some of these issues are discussed in this section.

1. Security
Security is the key to reconstruction and rebuilding program in the country. One of the few safe spaces in the entire country is the city of Kabul, under the protection of International security Assistance force (ISAF). Kabul is host to several military installations in Afghanistan. Military establishments like ISAF compound, the Bagram air base and various military camps have had an impact on the urban development of the city.

2. Donor funded projects
All the donor funded agencies look at Kabul as their port of entry for the beginning of their operations and have their establishments like offices, guesthouses, embassies etc. in Kabul. Kabul has the largest number of NGO having a base and operating within the city (Mitchell, 2017). The security situation in many parts of the country is fragile and therefore most International agencies are operating from Kabul and have not ventured outside the capital city. The central part of Kabul houses embassies, Government offices and Guesthouses of the various aid organizations. The central part of the city is also militarized zone with armed check points.

3. Migration
Afghanistan is historically situated at an important trade route and has always had movement of people within and outside its boundaries. Over the last 30 years, cross border movements and internal displacements have increased considerably due to the geo-political instability and severe droughts during the 1990's and 2000's (*Migration and Urban Development in Kabul: Classification or Accommodation*, 2012). After the fall of Taliban in October 2001 and the subsequent pouring in of International aid boosting the economy of the country, Kabul has experienced a high influx of refugee returnees from Pakistan and Iran. These refugees preferred settling in Kabul rather than their place of origin due to security reasons and lack of livelihood. A large number of Internally displaced people fleeing from conflict and drought prone provinces of Badghiz, Farah, Ghor and Herat have settled in Kabul. This has led to immense pressure on housing and other amenities for the growing population. Most of the refugees live in the remote areas on hillside in Kabul without access to basic services like water and electricity. This growth of the city and patterns of development need to be studied for planners, decision makers and policy makers. This research is an attempt to bring out the urban growth patterns and the factors influencing them in a conflict zone.

To understand urban growth patterns in a conflict zone the most reliable method is the use of satellite images and aerial photographs. The analysis of satellite image and aerial photographs has proven to be of use in complex civil war situations. They help gathering information about areas that are not safe to enter without security. Moreover, for effective urban planning and management it is essential to utilize advance techniques to obtain information on the patterns, state, characteristics of urban development. The research would involve the use of earth observation-based data, statistical information like census data and publications from the Govt. Authorities / Ministries.

3.2. Data

The research aimed at studying the change between periods 2001 (start of the conflict "operation enduring freedom") to 2017 for that efforts were made to acquire imagery prior to October 2001. There are no high-resolution imageries available with the various organization in Kabul shown in Table 1. The Built-up layer is downloaded from the Global Human Settlement Layer website which is not accurate because a lot of built-up has not been captured in the layer. Data is then acquired from consisting of built-up layer which was specifically extracted from IKONOS image for the research purpose only.

Aerial photographs with sub meter resolution for the year 2017 is acquired from CRIDA (Capital Region Independent Development Authority). Built-up data of 1 mts. resolution for the year 2001 extracted from IKONOS is acquired from JRC produced by MASADA software which was developed for generating the Global Human Settlement Layer. The aerial photograph for the year 2017 is processed to analyse patterns of growth in the city in a span ranging from 2001 to 2017. The data pre-processing requires resampling, mosaicking and subset of the scenes on the basis of the extent of study area which is the city of Kabul covering an area of 1023 km^2. Since the aerial photographs consists of noise, so one of the important steps in pre-processing involves removing the noise from the aerial photograph. The aerial photograph is classified using Random Forest algorithm in SAGA GIS. Spatial logistic regression is used for identifying factors impacting the growth patterns. The dimensions and the associated explanatory factors that are considered for logistic regression are mentioned in Table 6. The data is acquired from various secondary sources. Data acquired and the source of data is shown in Table 1.

Table 1 Data collected and its sources

Raster Data			
Years	Sensor	Resolution	Source
2017 MSS	Aerial photographs (3 bands R,G,B)	0.4 meters	CRIDA
2001 Built-up map	Built-up data (IKONOS)	1 meter	JRC
DEM		5 mts	CRIDA

Vector Data	
Shapefiles	Source
22 District Kabul city	AGCHO
Road network	Openstreetmap extracts https://extract.bbbike.org/
Population	CSO
Educational Institutions	AGCHO
Hotels and Restaurants	AGCHO
Embassies	AGCHO
Ministries	AGCHO

Hospitals	AGCHO
Religious places	AGCHO
Planned and unplanned settlement boundaries	CRIDA
Military Establishments	CRIDA
Locations attacked in last 10 years	Wikipedia
AGCHO: Afghan Geodesy and Cartography office, CRIDA: Capital Region Independent Development Authority, CSO: Central Statistical Organization, GHSL: Global Human Settlement Layer	

3.3. Methodology

The various steps of methodology are based on physical data collection from different sources in Kabul government publications, data provided by JRC (built-up layer extracted from IKONOS imagery of 1 meter resolution. The imageries procured from CRIDA were of poor radiometric quality the best imagery was chosen for further processing. The change in the planned and unplanned settlements is analyzed using high resolution aerial photograph provided by CRIDA and Built-up layer provided by JRC were used for processing. The methodology adopted for the model is shown in the flowchart Figure 4

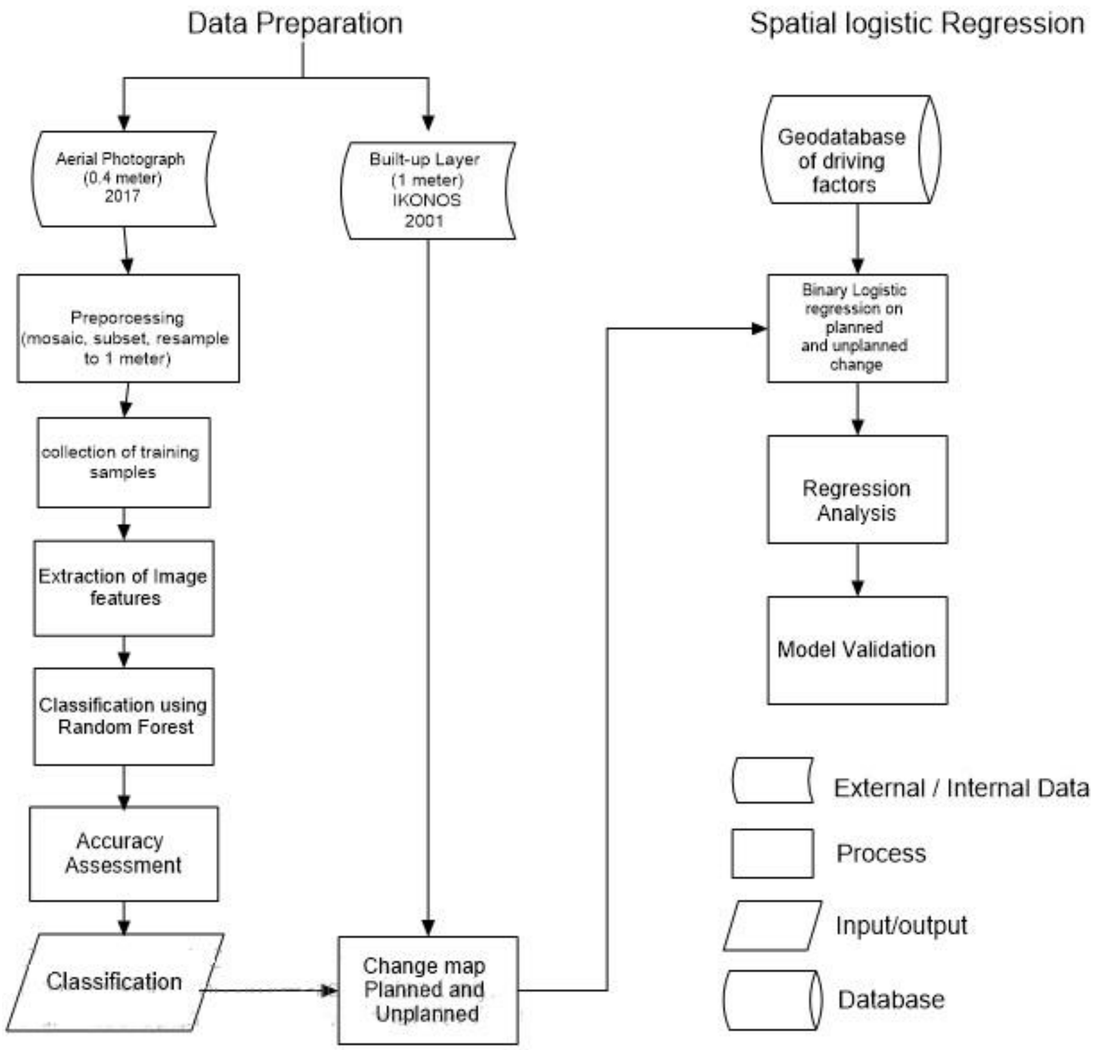


Figure 4 Flowchart showing the methodology

3.4. Classification

Aerial photographs for the year 2017 and the IKONOS built-up layer are of different spatial resolution therefore pre-processing of data involves resampling the aerial photograph to a resolution of 1 meter and mosaicking the scenes that cover the extent of the study area which is carried out in ERDAS Imagine. The photograph for the year 2017 consisted of noise and had pixels with no value which is taken care of by applying the masking option in ENVI. The pixels with no values were replaced by a value of 1. Random Forest classification is performed using SAGA GIS.

There are 2 main categories of classes the built-up and non-built-up. They are divided into planned and unplanned. The built-up classes are categorized on the basis of built-up densities. The building footprints are downloaded from OSM (open street map) website to be used to estimate the built-up densities of the settlement in the planned and unplanned parts of the city. A sample area size of 250 mts X 250 mts is used to identify the built-up classes on the basis of built-up densities. The cut off limits are set for the different densities. The cut off limits are 0-20 % as very low density, 20-40% as low density, 40 -60 % as medium density and above 60 - 80% as high density and above 90% as very high-density class. The classes were on the basis of delineated boundaries of planned and unplanned provided by CRIDA (Capital Region Independent development authority). In the case of Kabul, the built-up densities range from medium to low. There are mainly detached houses and very few high-rise buildings. Since the early 2000 few residential townships have been built and few more are under way to cater for the growing population of Kabul. These townships are mostly located on the outskirts of the city. All the hillside settlements built on slope are unplanned. However, the construction of the unplanned settlements is similar to the planned settlements. The only defining feature for the unplanned settlement is the lanes (less than 5 mts.) separating the buildings which is much narrower than the lanes in the planned settlements. The samples of the classes are shown in Table 2.

Table 2 Classification of Built-up and Non-Built-Up

Land-Use	Classes	Sub Classes	Sample set
Built-Up	Planned	Medium density 46% Built-up coverage	
		Low Density 25% Built-up coverage	

	Unplanned	Hillside settlement 20 % Built-up coverage	
		Medium density 40% built-up coverage	
Non – Built Up	Water		
	Bare Land		
	Vegetation		

3.4.1. Random Forest Classification

Random forest is chosen for classification as it can handle large number of features without affecting the overall accuracy of the classification. The computation time for processing several bands is greatly reduced by employing machine learning algorithms like Random forest (Ruiz Hernandez & Shi, 2018). Random Forest is supervised learning algorithm that can be used for both classification and regression tasks. Random forest belongs to the category of decision tree classification algorithm. Random forest classifier creates a set of decision trees from randomly selected subset of training set. It then aggregates the votes from different decision trees to decide the final class of the test objects (Bürgmann Tatjana, 2015). There are two parameters: the number of variables (M) in the random subset at each node and the number of trees (T) in the forest. The selection of parameter M has influence on the final error rate. If M is increased, both the correlation between the trees and the strength (classification accuracy) of individual tree in the forest are increased (Guan, Yu, Li, & Luo, 2012). M is the square root of number of features. The number of tress T can be as many as possible. However, at a certain point the increase in the number of trees results in high computation time and does not improve the prediction accuracy.

Advantages of using Random Forest algorithm for classification.

- Random forest is robust algorithm
- It incorporates spectral bands and other feature selection layers like soil index, water index, NDVI
- It incorporates texture feature for classification which include metrics like entropy, variance, morphology, line feature etc.
- It avoids overfitting
- Can deal with large number of features
- Helps with feature selection based on importance of variable
- It is user friendly because it only deals with 2 free parameters

  -number of trees
  -variables randomly selected as candidate at each split

Random Forest classification of aerial photograph is performed using the ViGrA tool in SAGA GIS software. There are two types of data required in supervised Random forest classification; variables and training data set. The variables used for the classification are shown in Table 3.

Table 3List of image features included in the Random Forest Classification

Image Feature	Description	Variable Type	Description
Spectral features	Spectral features provide information regarding the spectral response of objects, which differ for land coverage types, states of vegetation, soil composition, building materials (Duque et al., 2017)	VARI (Visible Atmospherically resistant Index)	VARI is a vegetation index that was designed to emphasize vegetation in the visible portion of the spectrum, while mitigating illumination differences and atmospheric effects. It is ideal for RGB and color images; it utilizes all three bands (Gitelson, Kaufman, Stark, & Rundquist, 2002).
		rgNDVI (Red green normalize difference vegetation index)	Where Visible red and Visible green are the spectral measurements of the red and green visible wavelengths, respectively (Graesser et al., 2012).
		rbNDVI (Red blue normalize difference vegetation index	Where Visible red and Visible blue are the spectral measurements of the red and blue visible wavelengths, respectively (Graesser et al., 2012).
Texture features	Textural features characterize the spatial distribution of intensity values of an image and provide information about contrast, uniformity, rugosity, etc (Duque et al., 2017)	GLCM (Grey level co-occurrence matrix)	GLCM is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image (Held & Committee, 1998).
Structural features	These features provide information regarding the spatial arrangement of elements in terms of the randomness or regularity of their distribution (Duque et al., 2017).	Edge detection filter	edge detection is a technique used to find the boundaries of features in an image. This uses an algorithm that searches for discontinuities in pixel brightness in an image that is converted to grayscale. ("Applying Edge Detection To Feature Extraction And Pixel Integrity," n.d.).

Vegetation Index is one of the spectral measures used for feature extraction and the most widely used is the Normalized Differential Vegetation Index (NDVI) which has been employed in classifying the urban forms. Urban neighbourhoods have similar spectral properties like rooftop and building material and the radiometric differences can be extracted by employing Vegetation indices (Graesser et al., 2012). Since the spectral range of the aerial photograph are limited to the visible part of spectrum therefore the visible bands are used to generate vegetation indices. Three different spectral indices have been computed in this study. The red-green normalized vegetation index (rgNDVI) and the red-blue normalized vegetation index (rbNDVI) and the visible atmospheric resistant index (VARI). The equations for the indices are given in Table 4.

Table 4 List of spectral Indices

Vegetation Indices	Equation
VARI (Gitelson et al., 2002)	(Green – Red) / (Green + Red -Blue)
rgNDVI (Graesser et al., 2012)	(Red – Green) / (Red + Green)
rbNDVI (Graesser et al., 2012)	(Red – Blue) / (Red + Blue)

The most commonly used texture measure is the Grey level co-occurrence matrix (GLCM). It is a statistical method used to determine texture that considers the spatial relationship of pixels. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM[1], and then extracting statistical measures from this matrix. The texture measure derived from GLCM are in the form of descriptive statistics like mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment within a given window size and shift. These textural statistics can be classified into three main groups: 1) the contrast group, which measures local variations inside a kernel (e.g., contrast); 2) the orderliness group, which measures the regularity versus disorder of pixel values (e.g., entropy); and 3) the statistic group, which computes basic descriptive statistics within the kernel (e.g., variance) (Kuffer et al., 2016). There are several texture measures like landscape metrics (patch analysis), wavelets, semivariograms, fractal analysis, and many other. Despite these complications, in terms of software accessibility, ease of use, and even intuitive grasp of the texture measures, GLCM remains a primary go-to tool, mainly because it is able to measure roughness, coarseness and directionality in one calculation (Held & Committee, 1998).

Texture measurement requires the choice of
• window size,
• direction of offset,
• offset distance,
• which channel to run,
• which measure to use.

The window size defines the area of samples used for GLCM tabulations and texture calculations. The length and height of window is restricted to an odd number in the range of 3 to 999 for example 3x3, 5x5, 7x7, 9x9 etc. The number of grey levels represents a range of sample intensity values. The number of grey levels determines the size of the GLCM. The minimum grey level is 0. The maximum grey level depends on the digitisation depth of the image. For an 8-bit-deep image it is 255. In a binary image a pixel can only take on either the value 0 or the value 255. In contrast, in a greyscale or colour image a pixel can take on any value between 0 and 255. The direction and distance define the spatial relationship between the reference pixel and the neighbour pixel. The typical orientation values are 00, 450, 900 and 1350 and typical distance values d=1(00, 900) and $\sqrt{2}$ (450, 1350). The choice of channels depends on the type of data used to generate GLCM. Some texture would be better in certain channels than others. For example, vegetation would be better in near infrared or red channel or a combination of them like vegetation indices. There are very large number of texture measures and many of the texture measures are correlated with one another. There are at most 4 to 5 truly Independent textures(“GLCM Texture Feature,” n.d.).

[1] https://www.mathworks.com/help/images/texture-analysis-using-the-gray-level-co-occurrence-matrix-glcm.html

One of the commonly used structural feature are the use of edge detection filters. Detection of edges in an image is a very important feature-extraction method and has been widely used in many computer vision and image processing applications ("Applying Edge Detection To Feature Extraction And Pixel Integrity," n.d.). Directional, or edge detection filters are designed to highlight linear features, such as roads or field boundaries. Linear filters use moving window on the image as a 3 x 3 matrix form. Classical method of liner filter to detect edges or linear features is convolution filter. It is calculated in the spatial domain as the weighted sum of pixels within the moving window (kernel) (Image, Analysis, & Pca, n.d.).

There are various edge detection filters like SOBEL, Laplacian, Roberts, Gaussian. edge detection t can be grouped into a) Gradient and b) Laplacian. The gradient method detects the edges by looking for the maximum and minimum in the first derivative of the image. while the Laplacian method searches for zero crossings in the second derivative of the image to find edges (Maini & Dr, n.d.). In the classification the Laplacian edge detection filter is chosen as structure feature for classification.

3.5. Spatial Logistic Regression

There are several statistical analysis techniques like multiple regression, linear regression, log linear and logistic regression. Compared with multiple regression and log-linear regression which analyse data where both independent and dependent variable is categorical or nominal while logistic regression is appropriate in estimating binary dependent variables. Urban growth phenomenon is a complex phenomenon which does not usually follow normal assumptions. Its factors are mostly a mixture of continuous and categorical variables (Cheng & Masser, 2003)

In the present study Binary logistic regression will be performed for analysing the factors influencing the patterns of urban growth in a city under conflict. The classified image and the IKONOS built-up layer is used to extract the change. There are 5 explanatory variables which are considered for the logistic regression. Different explanatory variables that were used for the spatial logistic regression are shown in Table 5.

Table 5 List of factors influencing patterns of urban growth

Dimension	Variables	Description	Rationale	Data Source	Data Type
Conflict	Vulnerable locations (Embassies, Ministries, hotels and commercial hubs) Locations of security basis	According to (Gov.UK, n.d.) The locations which are mostly likely to be attacked are places where large public crowds can gather. Hotels used by the government of Afghanistan and western nationals, ministries, military establishments, airports (including Kabul International) and religious sites public places frequented by	vulnerable locations / high profile locations and locations which have been attacked are least likely to be developed further as they are restricted by security and are considered unsafe for development	CRIDA and AGCHO	Vector layer derived from plotting points of vulnerable locations, bombed locations and army bases

		foreigners, including hotels, restaurants, shops and market places, wedding halls			
Socio-economic and cultural	Religious places	social and economic factors influence urban development. They attract and influence the direction of growth These indicators include civic amenities, infrastructure recreation etc. (Cheng & Masser, 2003)	Locations closer to socio-economic activities are more likely to be developed	AGCHO,	Layer with location of mosques, Embassies/ Ministries layer, wedding hall and restaurants layer, schools and hospital layer
	Location of Educational Institutions,				
	Location of Hospitals				
Accessibility	Accessibility to main roads	Population and urban growth are faster where infrastructure is made available. Locational advantages and mobility are important for urban growth (Cheng & Masser, 2003)	Easy access to social and economic activities like schools, hospitals, restaurants such locations have higher chances of development	AGCHO,	Road network layer, schools, hospitals, mosque layer, city centre layer
	Accessibility to community services (mosques, schools, hospitals				
	Accessibility to city centre (CBD)				
Demographic	Population density	Driven by population increase, urban growth alters the community's social, political and economic institutions with changing land use and also affects the local ecology and environment. (Aithal & Ramachandra, 2016)	Population drives urban development. Areas with high density population will also attract more socio-economic activities.	CSO,	Population layer
Physical Factors	Slope	Suitability of a location to develop can be impacted by bio-physical factors, for instance, slope layer needs to be taken into consideration in urban expansion model (Mahamud et al., 2016)	Steeper and elevated areas are less likely to be developed due to high cost of construction and higher risk of land instability	CRIDA	DEM layer

The above factors are considered keeping in mind the conflict situation in the country and the growth of the population from 2001 to 2017. These factors are unique to the city of Kabul. The socio-economic factors and the conflict factors are vulnerable locations since these locations have been attacked and have a probability of being attacked in the future. These vulnerable locations are embassies, ministries, mosques, educational institutions, hospitals, hotels and restaurants. All these factors are mostly centrally located and therefore have high locational similarity. The vulnerable locations are shown in the map in the Figure 5.

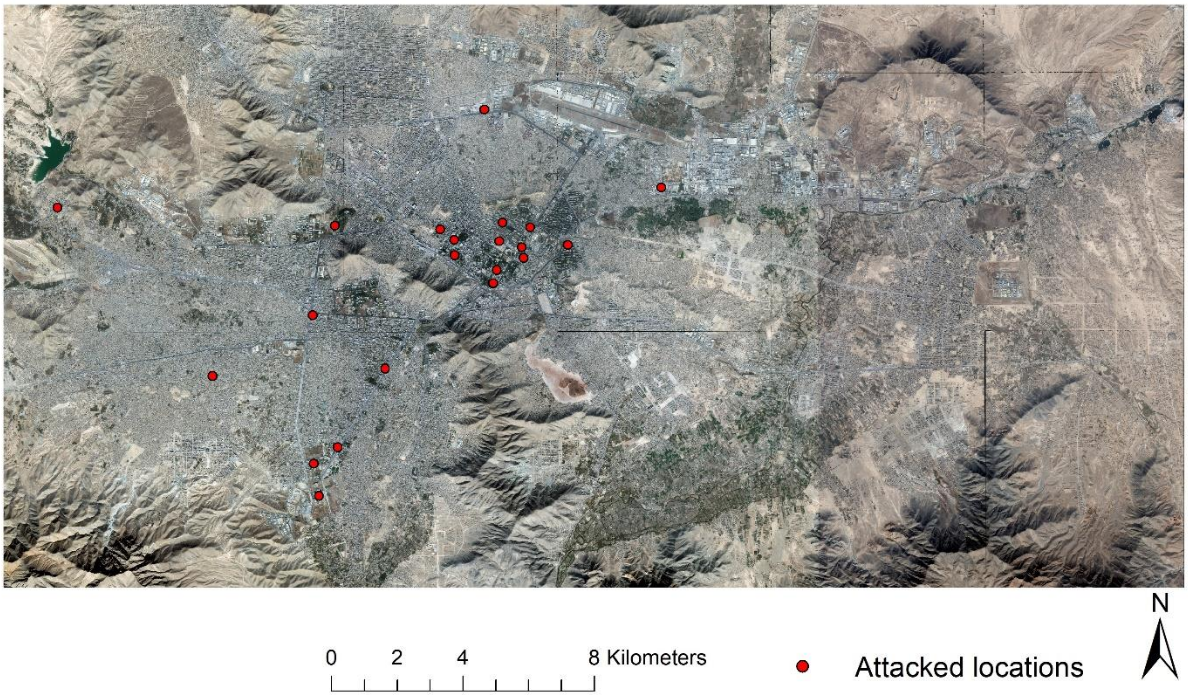


Figure 5 Attacked locations

4. RESULTS AND DISCUSSION

4.1. Classification Results

Classification of the aerial photograph is performed using Random Forest algorithm in SAGA GIS software. The classification involves 3 steps.

1) pre-processing of the aerial photograph which included mosaicking the scenes, resampling them to a resolution of 1 meter and subset to the extent of the study area and the removal of noise.
2) feature extraction which includes GLCM, rgNDVI, rbNDVI, VARI, Laplacian edge detection filter. DEM (Digital Elevation Model) and slope are additional geospatial features used as image features. There are total of 35 images features which are selected to be included in the Random Forest classification of the study area. The results of the feature importance table with permutation Importance and GINI decrease is shown in Figure 8 and 9. The parameter defined for the GLCM texture feature extraction are 1) processing window size of 3x3 as the resolution of the photographs was 1 meter the window size is best suited, 2) grey quantization levels set to 64 and 3) a co-occurrence shift of 1. The texture computed are mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment and correlation.
3) Collection of training samples. The training samples are collected using visual interpretation and equal number of samples are collected for each class (with a minimum point spacing of 5mts). As a proportion of the full image to be analysed the number of training samples would represent less than 1% to 5% (Guan et al., 2012). For accuracy assessment adequate number of reference samples for each class are collected from high resolution aerial photographs. According to Guan et al., (2012) rule-of-thumb of a minimum of 50 samples per class should be collected for a satisfactory confusion matrix. The number of training samples collected are 600 for each of the seven classes shown in the Figure 6.

References samples are collected from high resolution aerial photograph from the year 2016. For each class 50 samples were collected to check the accuracy of the classification. The collected samples are shown in Fig 7.

4.1.1. Feature Selection

One of the outputs of the classification is the feature importance table which measures the importance of the features by calculating the increase in the model's prediction error after permuting the features.
There are two measure of feature importance that can be used for image feature selection in Random forest classification.

a) GINI decrease: During the implementation of Random Forest, predictor variable importance is calculated by using GINI importance which is used to determine the nodes of individual decision trees to generate mean decrease of impurity (MDI) value, which is used to rank variables importance to the model. Higher mean decrease in GINI indicates higher variable importance (Altmann, Toloşi, Sander, & Lengauer, 2010)
b) Permutation Importance (PIMP): The PIMP of a feature is computed as the average decrease in model accuracy on the out of the bag samples when the values of respective features are randomly permuted (Altmann et al., 2010). The graphs Fig 7 and 8 show the feature importance results with the permutation importance and GINI decrease values of all the 35 image features used in the classification.

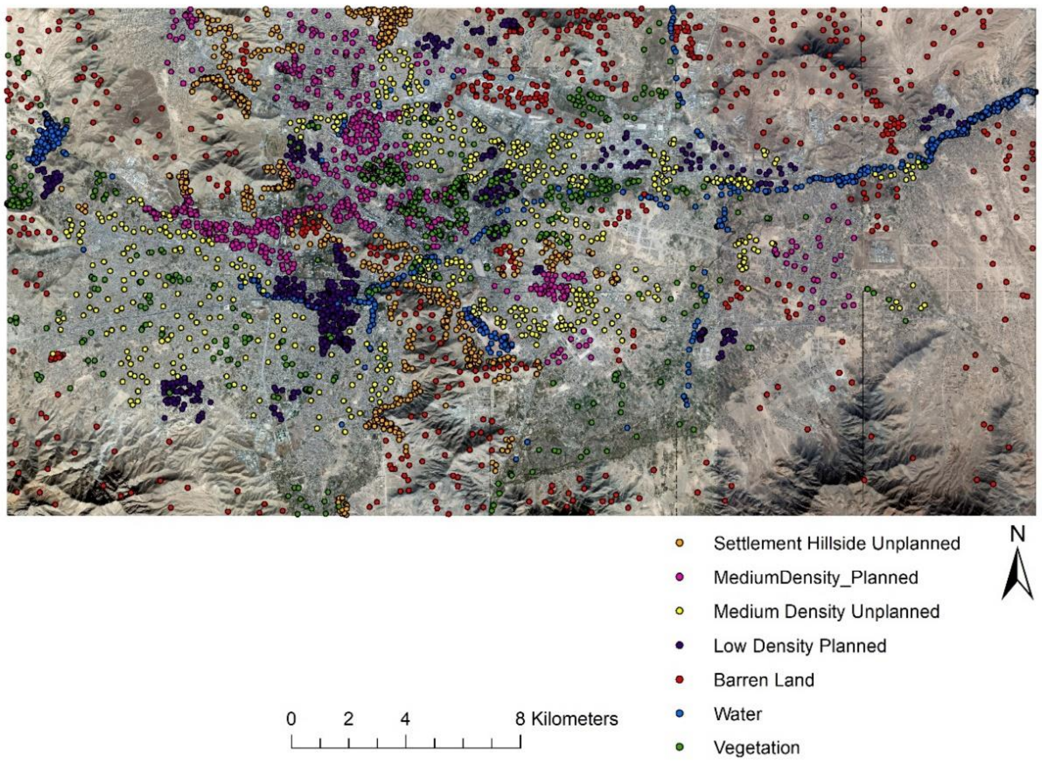


Figure 6 Training samples

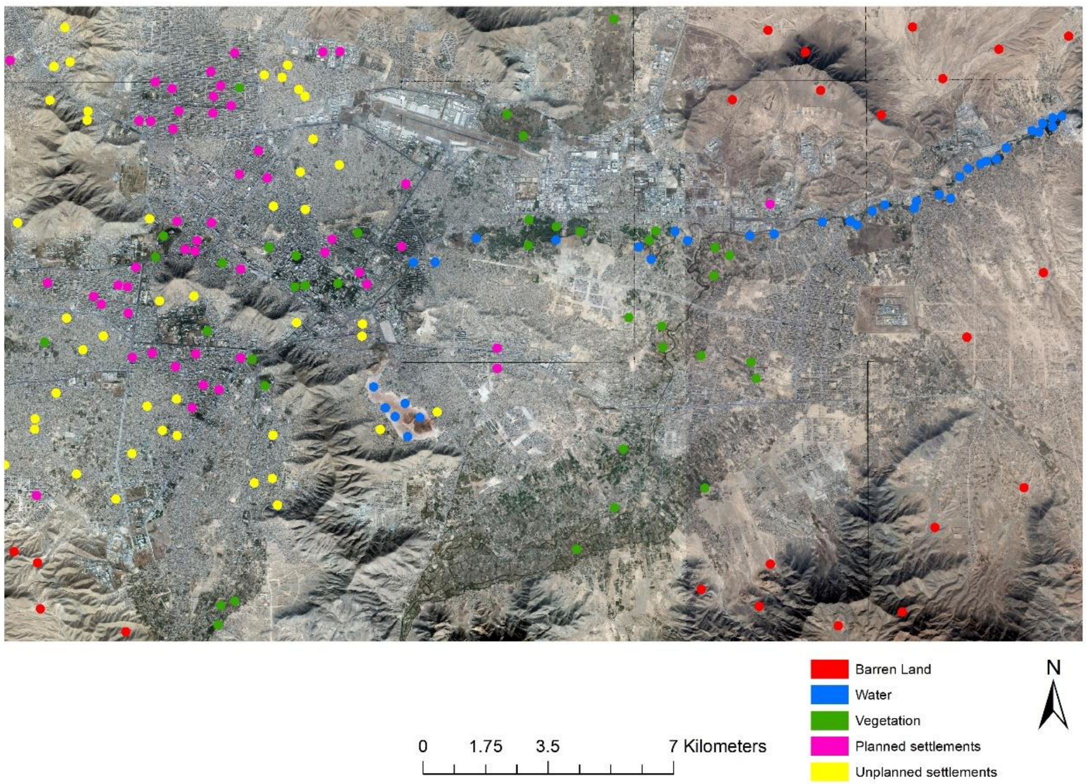


Figure 7 Reference samples

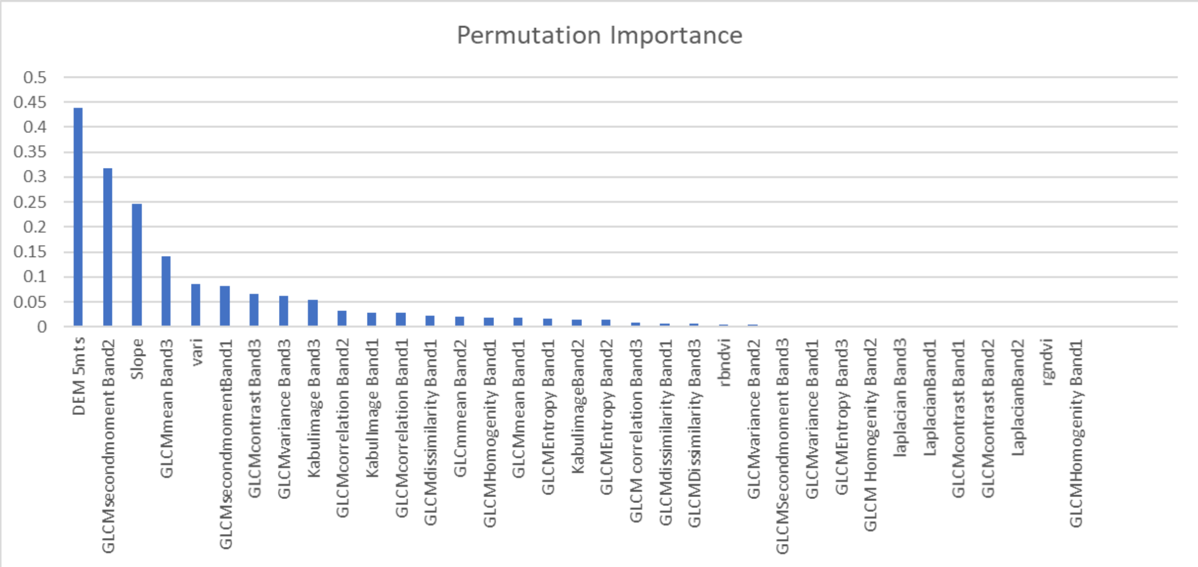


Figure 8 Permutation Importance

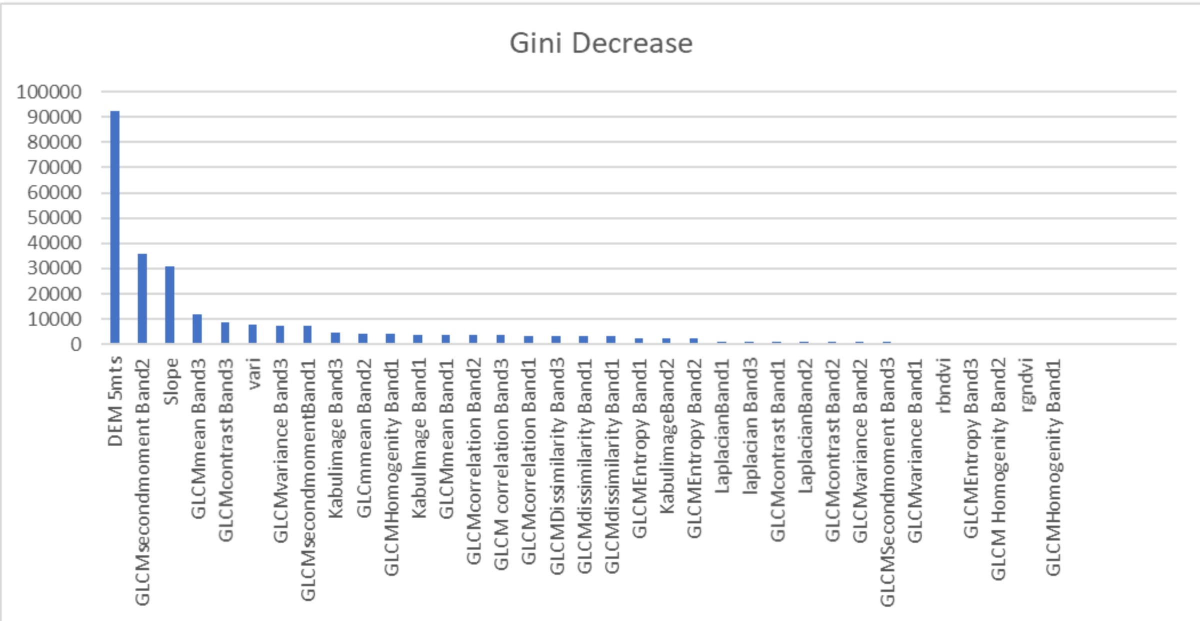


Figure 9 GINI Decrease

The classification is performed using all the 35 image features. As can be seen from the graphs Fig 8 and 9, DEM (Digital Elevation Model) and slope have high permutation importance and GINI decrease values. In the category of texture features the features with high permutation importance and GINI decrease are GLCM second moment Band2 and Band 1, GLCM mean Band 3. Among the spectral feature VARI (Visible Atmospheric Resistance Index) has a high permutation importance and GINI decrease. The original bands 1 and 3 of the aerial photographs have a high permutation importance and GINI decrease. The output of the classification in shown in Figure 10.

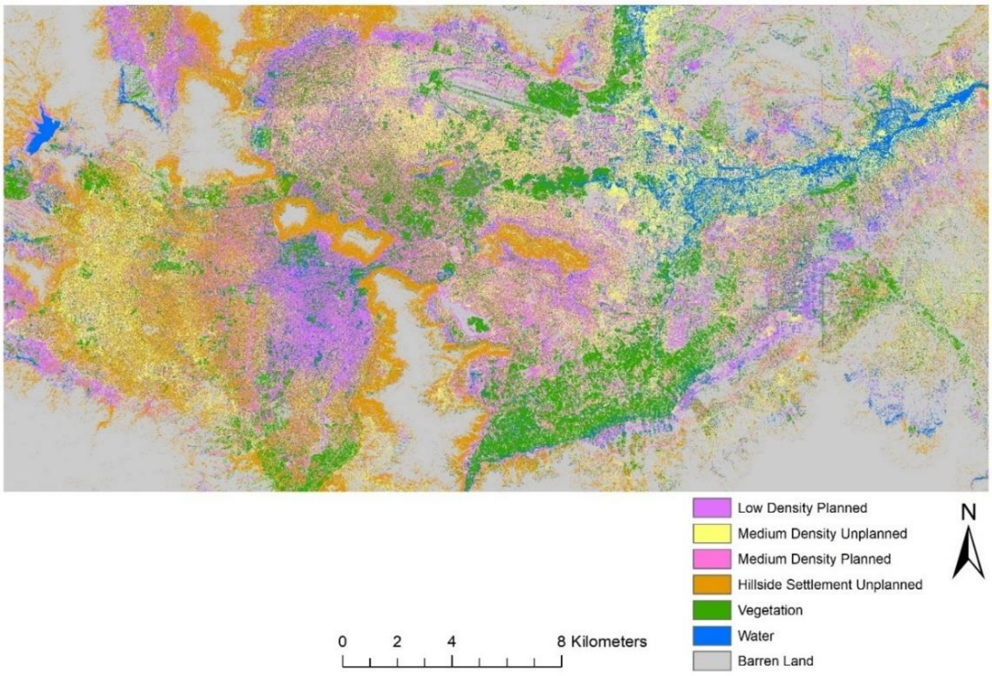


Figure 10 Result of Random Forest Classification with 7 classes

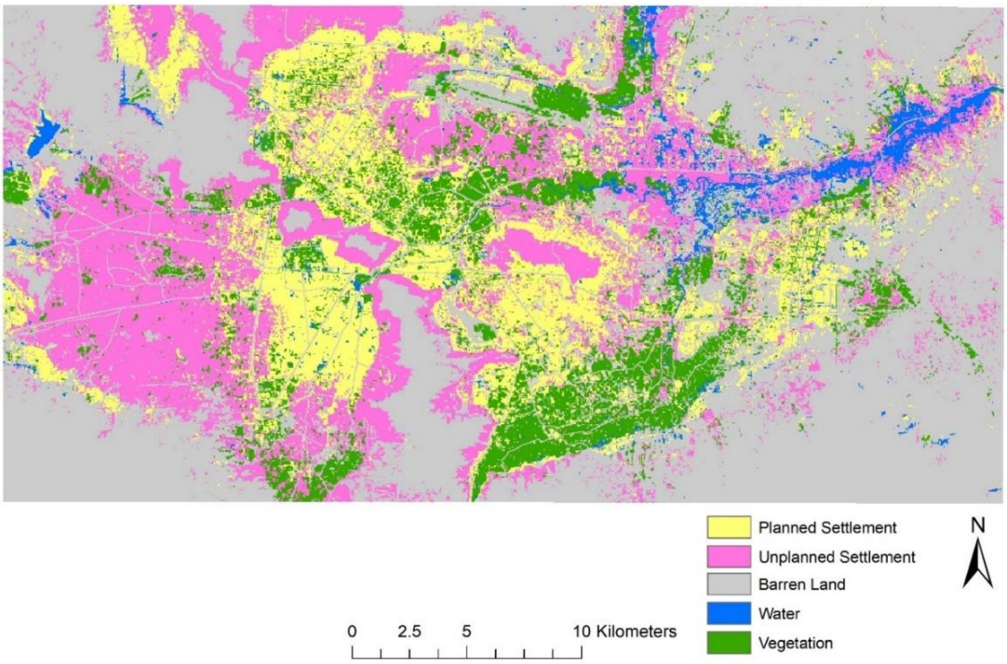


Figure 11 Results of the classification with 5 classes

The overall accuracy achieved for the classification with seven classes is 49% as shown in Table 6. The lowest accuracies achieved are for medium density planned and low density planned as can be seen in Table 7. The user accuracy for low density is 29 % and producer accuracy is 15% and the user accuracy of medium density planned is 39% and producer accuracy is 22 %.,

Table 6 Summary of the classification accuracy with seven classes

Name	Value
Kappa	0.41324
Overall Accuracy	0.497001

Table 7 User and producer accuracy of Individual classes

CLASS	Barren	Hillside Unplanned	Medium Den Planned	Medium Den Unplanned	Low Density Planned	Water	Vegetation	SumUser	AccUser
Barren	5861	1145	1178	1772	3293	1401	830	15480	37.861757
Hillside Settlement Unplanned	347	2698	1507	1268	1373	101	59	7353	36.692506
Medium density Planned	272	898	1742	848	582	74	8	4424	39.37613
Medium Density Unplanned	538	1930	1889	3095	1102	26	42	8622	35.896544
Low Density Planned	105	945	1173	616	1185	6	3	4033	29.382594
Water	65	14	30	8	3	5900	20	6040	97.682119
Vegetation	665	233	342	246	306	360	6860	9012	76.120728
SumProd	7864	7863	7861	7853	7846	7868	7857		
SumAcc	74.5295	34.312603	22.160031	39.41169	15.103237	74.98729	87.310678		

The low accuracy of the classification will affect the accuracy of the regression model therefore to improve the classification accuracy and to perform the logistic regression the built-up was combined as planned and unplanned by merging the 2 planned classes (Medium Density planned, and low density planned) and the 2 unplanned classes (Hillside Settlement and Medium Density Unplanned). The accuracy assessment is performed for the merged classes. Reference samples are collected roads are removed from the classification as they are not included in Built-up layer of 2001. The result of the merging of the built-up classes is shown in Fig 11.

4.2. Accuracy Assessment

The validation of the classification for the aerial photograph is based on the reference points collected for the 5 classes using high resolution aerial photograph of 2016 as the basis. The points are compared with the classification using the inbuilt confusion matrix (polygon / Grid) in SAGS GIS software.

Table 8 Summary of Classification Accuracy

NAME	VALUE
Kappa	0.51639
Overall Accuracy	0.613116

The overall accuracy (seen in Table 8) of the classification is 61% which means 61% of the pixels were correctly assigned. The Kappa coefficient is another indicator for checking the accuracy. It takes values from 0 to 1. If the coefficient is equal to 0 then there is no agreement between the classification and referenced samples and if it is equal to 1 then the classified image and reference completely match.

Table 9 User and producer accuracy for individual classes

CLASS	Unplanned	Planned	Vegetation	Barren Land	Water	SumUser	AccUser
Unplanned	8970	3146	1791	1130	1328	16365	54.812099
Planned	3869	6880	957	582	576	12864	53.482587
Vegetation	1330	3642	7842	250	589	13653	57.437926
Barren	959	971	3142	13155	2035	20262	64.924489
Water	506	1015	1898	536	11095	15050	73.72093
SumProd	15634	15654	15630	15653	15623		
AccProd	57.374952	43.950428	50.172745	84.041398	71.01709		

The individual class accuracy can be seen in Table 9. The lowest Producer accuracy is for class planned settlement which is 43% and vegetation which is 50%. The lowest user accuracy is for planned settlement with an accuracy of 53% and unplanned settlement with accuracy of 55%. The low accuracy for planned and unplanned is due to the low contrast between the two classes. The roof tops for both the planned and the unplanned settlement have similar texture. Highest accuracies are achieved for Barren land in case of producer accuracy with 84% and for the class water for user accuracy with 74%.

The errors of commission in the confusion matrix is a measure of producer's accuracy (AccProd) which is the highest for classes barren land 84% and water with 71 % followed by unplanned with 57% vegetation with 50% and planned with 43% accuracy. The errors of omission which is a measure of user accuracy (AccUser) is the highest for water which is 73% followed by Barren which is 65% as can be seen in the confusion matrix in Table 9.
To perform the logistic regression the accuracy of the built-up layer extracted from IKONOS is checked with the accuracy of the classification. The accuracy of the built-up layer extracted from IKONOS is 75% (provided by JRC) for the 3 classes i.e. built-up, non-built-up and water. The accuracy of the combined built-up class (planned and unplanned), water, vegetation and barren for the aerial photograph is 70 %. The accuracies of the two data sets is comparable and therefore are considered to perform regression. The result for the accuracy of the combined built-up class and the 3 non built-up classes is shown in Table 8 and the accuracy of the combined Built-up (planned and unplanned) and water is shown in Table 9.

Table 10 Accuracy results of classification with Built and Vegetation, barren and Water class

	Built	Vegetation	Barren Land	Water		
Built	22865	2748	1712	1904	29229	0.78
Vegetation	4972	7842	250	589	13653	0.57
Barren	1930	3142	13155	2035	20262	0.65
Water	1521	1898	536	11095	15050	0.74
	31288	15630	15653	15623	78194	
	0.73	0.50	0.84	0.71		**0.70**

Table 11 Accuracy result of the combined built-up and Water class

	Built	Non-built	Water	Sum	
Built	22865	4460	1904	29229	0.78
Non-built	6902	24389	2624	33915	0.72
Water	1521	2434	11095	15050	0.74
Sum	31288	31283	15623	78194	
	0.73	0.78	0.71		**0.75**

A comparison of the overall accuracy for three classes i.e. built-up, non-built-up and water for built-up layer from Global human settlement layer and the classification of the aerial photograph of 2017 is shown in the table 12.

Table 12 Comparison of the overall accuracy of Global Human Settlement layer and Classification

	Global Human Settlement Layer	Aerial Photograph Classification
Kappa	0.301561	0.51639
Overall Accuracy	0.759174	0.70

4.3. Results of the Logistic Regression

To perform the logistic regression built-up layer of Kabul for the year 2001 at a resolution of 1 meter is extracted from IKONOS acquired from Global Human Settlement layer (GHSL). The two images are then used to see the change in the built-up areas from the year 2001 to 2017. The change in the settlement from 2001 to 2017 is mapped in the form of a binary classification i.e. what was built in 2001 is 0 with no data and all that was built after 2001 is 1 that is the change in the built-up from 2001 to 2017. The change in the built-up forms the dependent variable as can be seen in Figure 12.

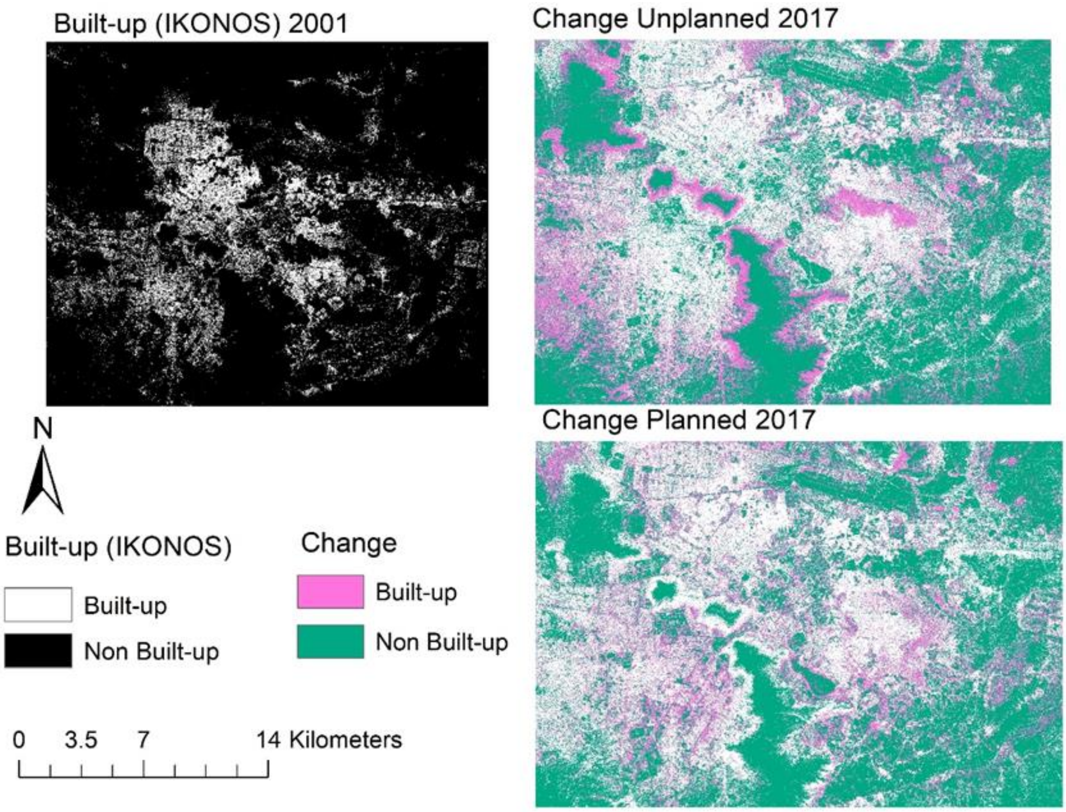


Figure 12 Changed in planned and unplanned settlement from 2001 to 2017

4.4. Amount and patterns of growth in Planned and Unplanned Settlements

The total built-up area calculated for the planned settlements in 2001 is 71 km^2 and in 2017 is 94 km^2. Thus, the growth in the planned areas from 2001 to 2017 is 22 km^2. The planned areas have grown 1.25 times from 2001 to 2017. The total built-up area calculated for unplanned settlements in 2001 is 21 km^2 and in 2017 is 114 km^2. Thus, the growth in unplanned from 2001 to 2017 is 94 km^2. The unplanned settlements have grown 4.5 times in period from 2001 to 2017 as seen in the Table 13. The tremendous growth in the unplanned settlement is peculiar to Kabul, being the capital of Afghanistan compared to the other cities having the largest presence of military bases, NATO base and Afghan troops in the whole country giving a feeling of security to the population apart from economic opportunities in the city. As can be seen from Figure 12 unplanned growth is mainly on the west, north west, south west and on the hillside. The settlements on the hillside are all unplanned. A large part of the planned growth is on central, east, north east and southern parts of the city. Therefore, from the results it can be said that the unplanned growth has occurred mainly in the western parts of the city whereas the planned growth is mainly toward central and eastern parts of the city. Overall it can be said that Kabul is a sprawling city because of low rise and medium density built-up. The increase in the unplanned sectors because of the conversion of large agricultural lands into built-up as well as land on the steeper slopes of the hills show the city sprawl. Most of the houses in the unplanned settlements are separated dwellings and have central courtyard and similar building material as in the planned settlements as is clearly visible on the aerial photograph which makes it difficult to distinguish.

Table 13: Growth in the planned and unplanned settlements

Urban Growth	Built-up in 2001 (in km ^{2})	Built-up in 2017 (in km^2.)	Growth (in km^2.)	Growth (in % age) 2001 to 2017
Planned settlements	71	94	22	30
Unplanned settlements	21	114	94	447
Total Built-up	92	208	116	126

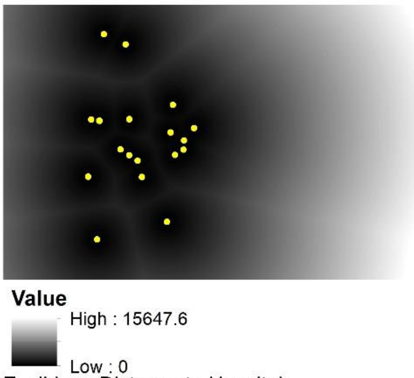
4.5. Binary Logistic Regression

Two binary logistic regression is performed on the change in the built-up in planned settlements and on the change in the built-up on the unplanned settlement. The regression model is used to study the importance of the predictor variables on the change in the built-up in the planned and unplanned settlements. The list of variables is shown in Table 14 and the predictor variables chosen for the regression are shown in Figure 13.

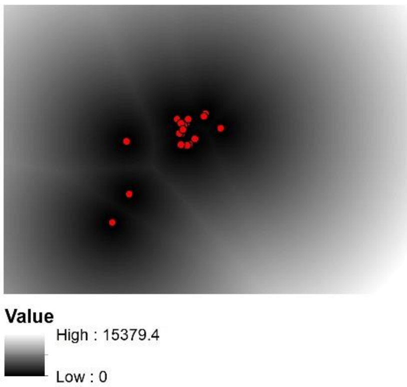
Table 14 List of Variables for the logistic regression

Variable	Definition	Unit	Type of Variable
Dependent			
Change in Built-up	0 no change in Built-up 1 Change in Built-up		dichotomous
Independent			
Major Roads	Euclidean distance	meters	continuous
Vulnerable Locations (Embassies and Ministries, Hospitals, Hotels, Mosques and Educational Institutions)	Euclidean distance	meters	continuous
Military bases	Euclidean distance	meters	continuous
Population density		km^2.	continuous
Slope		degrees	continuous

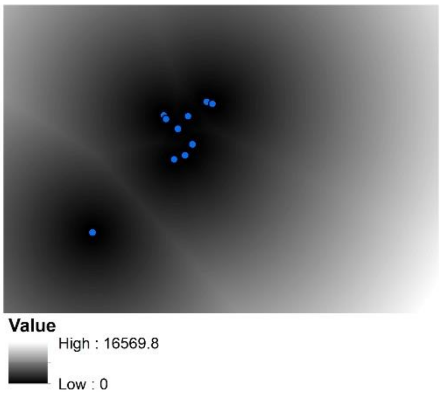
Euclidean Distance to Educatioanl Institutions



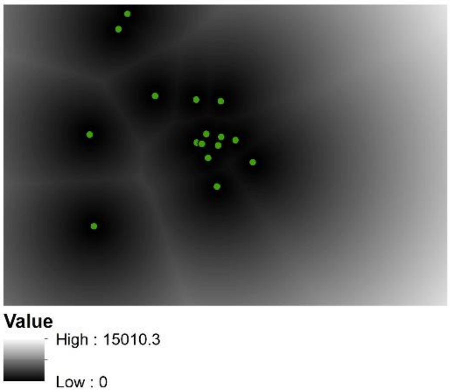
Euclidean Distance to Embassies and Ministries



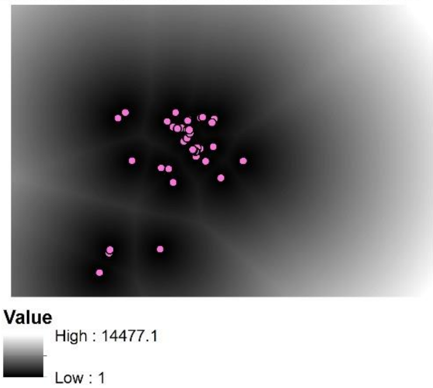
Euclidean Distance to Hospitals



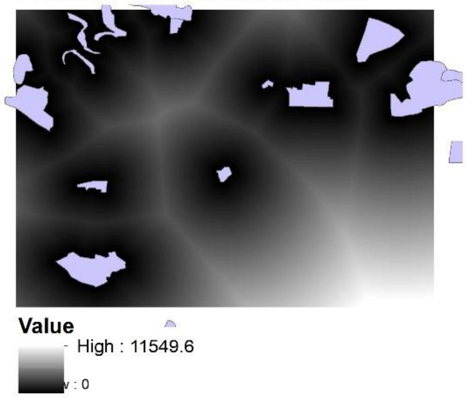
Euclidean Distance to Mosques



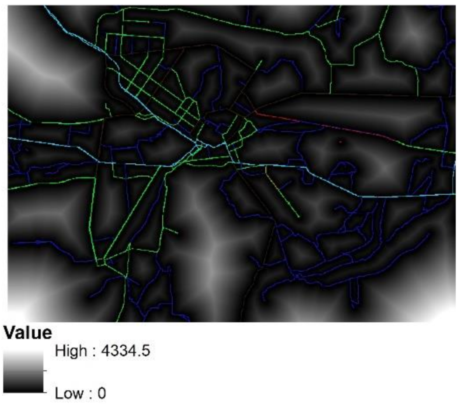
Euclidean Distance to Hotels and Restaurants



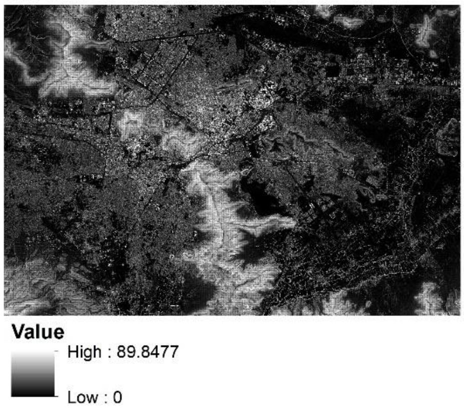
Euclidean Distance to Military Bases

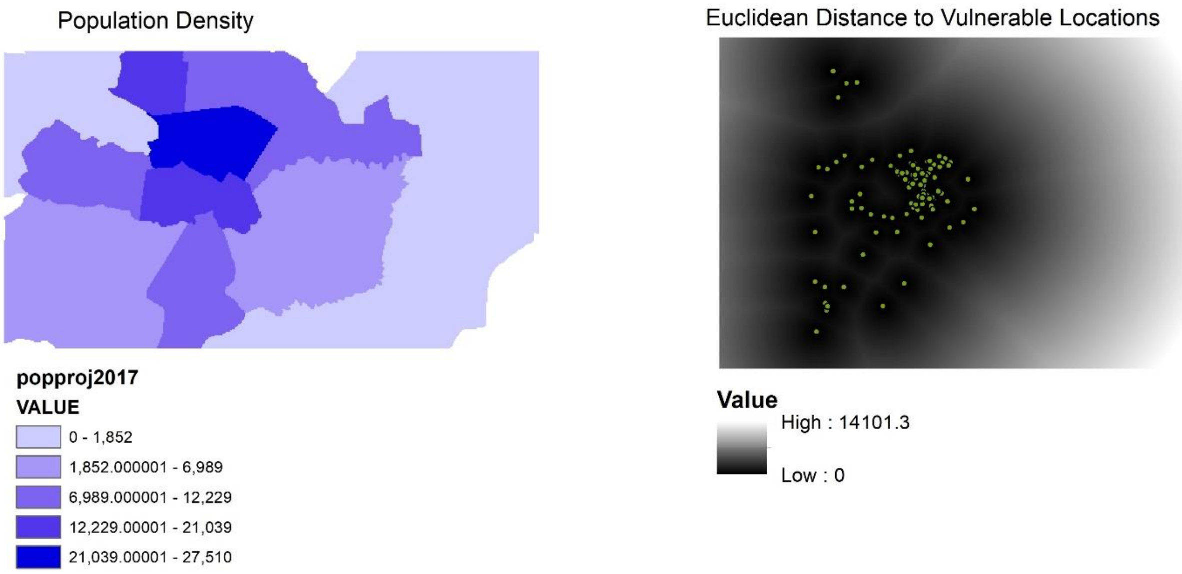


Euclidean Distance to Major Roads



Slope





*Vulnerable locations are combined Educational Institutions, Embassies and Ministries, Hospitals, Mosques, Hotels and Restaurants

Figure 13 Predictor variables for regression

Pearson's correlation coefficient is used to measure the strength of linear relationship between pairs of variables. The variable planned is a dichotomous variable having values of 1 as built-up and 0 as non-built-up shown in Table 14. The correlation ranges between -1 and +1. A correlation of +1 indicates a perfect positive correlation between the variables. A value in one variable is associated with an increase in value in another variable. A correlation of -1 indicates a perfect negative correlation which means an increase in value in one variable is associated with decrease in the other variable. A correlation coefficient of 0 indicates there is no liner relationship between the variables.

Table 15 Pearson's coefficient correlation for variables for change in planned settlements

Correlations

		Planned	Population Density	Slope	Education Institutions	Hospitals	Mosques	Military bases	Hotels	Embassies and Ministries	Roads
Planned	Pearson Correlation	1	.182**	-.052**	-.186**	-.201**	-.220**	-.106**	-.189**	-.218**	-.232**
	Sig. (2-tailed)		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	33254	32686	33254	33254	33254	33254	33254	33254	33254	33254
Population	Pearson Correlation	.182**	1	.119**	-.630**	-.674**	-.629**	-.245**	-.639**	-.668**	-.250**
	Sig. (2-tailed)	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	32686	50762	50762	50735	50735	50735	50735	50735	50735	50735
Slope	Pearson Correlation	-.052**	.119**	1	-.218**	-.139**	-.179**	.065**	-.177**	-.121**	.250**
	Sig. (2-tailed)	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	33254	50762	51445	51410	51410	51410	51410	51410	51410	51410
Education Institutions	Pearson Correlation	-.186**	-.630**	-.218**	1	.890**	.965**	.406**	.925**	.867**	.124**
	Sig. (2-tailed)	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000
	N	33254	50735	51410	51410	51410	51410	51410	51410	51410	51410
Hospitals	Pearson Correlation	-.201**	-.674**	-.139**	.890**	1	.892**	.416**	.941**	.975**	.248**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000
	N	33254	50735	51410	51410	51410	51410	51410	51410	51410	51410
Mosques	Pearson Correlation	-.220**	-.629**	-.179**	.965**	.892**	1	.354**	.935**	.880**	.173**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000
	N	33254	50735	51410	51410	51410	51410	51410	51410	51410	51410
Military bases	Pearson Correlation	-.106**	-.245**	.065**	.406**	.416**	.354**	1	.315**	.395**	.150**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000
	N	33254	50735	51410	51410	51410	51410	51410	51410	51410	51410
Hotels	Pearson Correlation	-.189**	-.639**	-.177**	.925**	.941**	.935**	.315**	1	.957**	.190**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000
	N	33254	50735	51410	51410	51410	51410	51410	51410	51410	51410
Embassy and Ministries	Pearson Correlation	-.218**	-.668**	-.121**	.867**	.975**	.880**	.395**	.957**	1	.288**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000
	N	33254	50735	51410	51410	51410	51410	51410	51410	51410	51410
Roads	Pearson Correlation	-.232**	-.250**	.250**	.124**	.248**	.173**	.150**	.190**	.288**	1
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	N	33254	50735	51410	51410	51410	51410	51410	51410	51410	51410

**. Correlation is significant at the 0.01 level (2-tailed).

There is a significant positive correlation between planned settlements and population density and negative correlation with mosques and roads. Population has a negative correlation with embassies, mosques, educational Institutions. Slope has a negative correlation with planned settlements and education. The variables Hotels, embassies and ministries, hospitals, educational institutions are high correlated as can be seen in Table 15.

The first regression is performed on the change in the planned settlements. Before performing the regression, we check for the collinearity of the factors. The collinearity of the predictor variables was checked using the SPSS software. The variables which are highly correlated will not be included for regression. To estimate the collinearity the variance inflation factor (VIF) is checked and as a rule of thumb if the VIF of a variable is greater than 10, it should not be included in the regression.

Table 16 Variance Inflation Factor for the predictor variables for change in planned settlements

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Educational Institution	0.046	21.792
	Hospital1	0.034	29.277
	Mosques	0.049	20.466
	Military bases	0.607	1.648
	Hotels	0.030	33.202
	EmbassyMin	0.023	43.436
	Roads	0.693	1.442
	Population Density	0.477	2.098
	Slope	0.801	1.248

a. Dependent Variable: Planned

As can be seen in the Table 16 Embassies and ministries, Hotels, hospitals, Educational Institutions and mosques have VIF of more than 10. These variables are centrally located and have locational similarities and are therefore combined into one variable called vulnerable locations. These locations have a high probability of being targets and have been attacked in the past therefore they are vulnerable locations.

Table 17 VIF after removing variables with very high values for change in planned settlements

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Population Density	0.541	1.848
	Slope	0.824	1.214
	Military bases	0.808	1.237
	Vulnerable Locations	0.488	2.049
	Roads	0.815	1.228

a. Dependent Variable: Planned

As can be seen from Table 17 all the variables have a VIF of less than 10 and therefore are included in regression

Table 18 Significance, Odds Ratio, Wald statistics results of the regression for planned settlements

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Population Density	0.039	0.003	152.158	1	0.000	1.040
	Slope	-0.003	0.001	6.539	1	0.011	0.997
	Roads	-4.775	0.137	1218.932	1	0.000	0.008
	Military bases	-0.180	0.077	5.456	1	0.020	0.835
	Vulnerable locations	-1.119	0.079	202.286	1	0.000	0.327
	Constant	-0.239	0.049	23.361	1	0.000	0.788

a. Variable(s) entered on step 1: Population, Slope, Roads1, Military1, Vulnerabl1.

All the factors are significant except slope and military bases as can be seen in Table 18. According to the Wald statistics the coefficients for roads, vulnerable locations and population density is significantly high. Exp(B) which is also called the odds ratio which is a measure to compare the relative importance of each factor. As can be seen from the table population density has the highest odds ratio followed by slope and roads has the lowest odds ratio. The negative signs for slope, roads, military bases and vulnerable location indicates inverse correlation as distance increase from these locations the probability of building in a cell decrease. Population density is positive factor which indicates that areas close to these locations have greater chance of getting developed. However, we must consider that the population density is calculated for a district but within the district some parts maybe sparsely populated. Areas with gradual slope have greater chances of getting developed while areas with steep slopes have less chance.

Table 19 Model summary for regression on change in planned settlements

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	34987.963[a]	0.103	0.149

a. Estimation terminated at iteration number 5

According to the model summary (seen in Table 19) the pseudo R square is 0.103 and 0.149. The classification table (seen in Table 20) indicate that the overall predicted percentage accuracy of the classification is 72%.

Table 20 Classification table results for change in planned settlements

Classification Table[a]					
Observed			Predicted		
			Planned		Percentage Correct
			0	1	
Step 1	Planned	0	23143	666	97.2
		1	8423	574	6.4
	Overall Percentage				72.3

a. The cut value is .500

The bivariate Pearson's correlation coefficient is used to measure the linear relationship between the pairs of variables. The variable unplanned is a dichotomous variable having values of 1 as built-up and 0 as non-built-up shown in Table 20.

Table 21 Pearson's coefficient correlation for variables for change in unplanned settlements

Correlations

		Unplanned	Population	Slope	Embassies and Ministries	Educational Institution	Hospital1	Roads	Mosques	Military bases	Hotels
Unplanned	Pearson Correlation	1	.143**	.035**	-.170**	-.195**	-.175**	-.164**	-.204**	-.143**	-.157**
	Sig. (2-tailed)		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	37311	37311	37311	37311	37311	37311	37311	37311	37311	37311
Population	Pearson Correlation	.143**	1	.091**	-.656**	-.628**	-.657**	-.255**	-.629**	-.249**	-.637**
	Sig. (2-tailed)	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	37311	51447	51447	51409	51409	51409	51409	51409	51409	51409
Slope	Pearson Correlation	.035**	.091**	1	-.124**	-.219**	-.142**	.242**	-.181**	.063**	-.179**
	Sig. (2-tailed)	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	37311	51447	51447	51409	51409	51409	51409	51409	51409	51409
Embassies and Ministries	Pearson Correlation	-.170**	-.656**	-.124**	1	.867**	.975**	.287**	.879**	.392**	.956**
	Sig. (2-tailed)	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000
	N	37311	51409	51409	51409	51409	51409	51409	51409	51409	51409
Educational Institution	Pearson Correlation	-.195**	-.628**	-.219**	.867**	1	.889**	.123**	.965**	.404**	.924**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000
	N	37311	51409	51409	51409	51409	51409	51409	51409	51409	51409
Hospital1	Pearson Correlation	-.175**	-.657**	-.142**	.975**	.889**	1	.247**	.892**	.414**	.941**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000
	N	37311	51409	51409	51409	51409	51409	51409	51409	51409	51409
Roads	Pearson Correlation	-.164**	-.255**	.242**	.287**	.123**	.247**	1	.171**	.149**	.189**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000
	N	37311	51409	51409	51409	51409	51409	51409	51409	51409	51409
Mosques	Pearson Correlation	-.204**	-.629**	-.181**	.879**	.965**	.892**	.171**	1	.352**	.935**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000
	N	37311	51409	51409	51409	51409	51409	51409	51409	51409	51409
Military bases	Pearson Correlation	-.143**	-.249**	.063**	.392**	.404**	.414**	.149**	.352**	1	.313**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000
	N	37311	51409	51409	51409	51409	51409	51409	51409	51409	51409
Hotels	Pearson Correlation	-.157**	-.637**	-.179**	.956**	.924**	.941**	.189**	.935**	.313**	1
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	N	37311	51409	51409	51409	51409	51409	51409	51409	51409	51409

**. Correlation is significant at the 0.01 level (2-tailed).

There is a positive correlation between unplanned and population density and slope. Roads have significant positive correlation with hospitals. Embassies and ministries, hotels, mosques, educational institutions and hospitals are highly correlated as can be seen in Table 21.

The binomial logistic regression is performed on the change in the unplanned areas. The collinearity was estimated for all the factors.

Table 22 Variance Inflation Factor for the predictor variables for change in unplanned settlements

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Population	0.539	1.856
	Slope	0.801	1.248
	Embassies and Ministries	0.023	43.720
	Educational Institution	0.044	22.685
	Hospital1	0.038	26.102
	Roads	0.727	1.375
	Mosques	0.052	19.387
	Military bases	0.612	1.635
	Hotels	0.033	30.729

a. Dependent Variable: Unplanned

The VIF for Embassies and ministries, Hotels, hospitals, Educational Institutions and mosques is more than 10 as can be seen from the Table 22. These variables are centrally located and have locational similarities and are therefore combined into one variable called vulnerable locations. These locations have a high probability of being targets and have been attacked in the past therefore they are vulnerable locations

Table 23 VIF after removing variables with very high values for change in unplanned settlements

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Population	0.567	1.764
	Slope	0.829	1.206
	Vulnerable Locations	0.501	1.996
	Military bases	0.800	1.249
	Roads	0.837	1.195

a. Dependent Variable: Unplanned

As can been seen from the Table 23 all the variables have a VIF of less than 10 and therefore have been included in the regression.

Table 24 Significance, Odds Ratio, Wald statistics results of the regression for unplanned settlements

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Population Density	0.026	0.003	91.140	1	0.000	1.026
	Slope	0.000	0.000	1.768	1	0.184	1.000
	Vulnerable Locations	-0.841	0.067	157.005	1	0.000	0.431
	Military bases	-0.761	0.065	138.054	1	0.000	0.467
	Roads	-2.422	0.092	695.660	1	0.000	0.089
	Constant	-0.006	0.041	0.023	1	0.880	0.994
a. Variable(s) entered on step 1: Population Density, Slope, Vulnerabl1, Military1, Roads1.							

All the factors are significant as seen in Table 24 except slope. The Wald statistics shows that Vulnerable locations, Military bases and roads coefficients are significantly high. The odds ratio i.e. Exp(B) is high for population density and slope and lowest for roads. The negative sign of the coefficient for vulnerable locations, military bases and roads indicate inverse correlation, as the distances increases from these locations the possibility of them being developed reduces. The factors of population density and slope have a positive coefficient which indicates that irrespective of the distance the possibility of areas closer to these locations getting developed is high. Population density and Slope are positive in the case of unplanned as there are more unplanned settlement on steeper slopes as compared to planned settlements.

Table 25 Model summary for regression on change in unplanned settlements

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	44959.786[a]	0.065	0.089
a. Estimation terminated at iteration number 5			

Looking at the model summary (seen in Table 25) the pseudo R square is 0.062 and 0.088. The classification result (seen in Table 26) indicates the overall predicted percentage accuracy of the classification is 65%.

Table 26 Classification table results for change in unplanned settlements

Classification Table[a]					
			Predicted		
			Unplanned		Percentage Correct
Observed			0	1	
Step 1	Unplanned	0	21852	1874	92.1
		1	10926	1918	14.9
	Overall Percentage				65.0
a. The cut value is .500					

The analysis of the regression model indicates that population density is most significant factors for the growth of both planned and unplanned settlements. In case of unplanned settlement slope is also one of the factors affecting the growth. It can be seen from the results that military bases also have a significant impact on the growth.

ROC curve is a tool used in evaluate and compare predictive models like logistic regression. The ROC curve is a graph that plots sensitivity (true positive) against 1-specificity (true negative). A model with high separation ability will have high sensitivity and specificity, which is achieved when the curve is closer to the top left corner of the plot. The graph in Figure 14 shows the ROC curve (Receiver operating characteristic) representing the variables used in the regression for planned change. The accuracy of the test depends on how well test separates the test variables which is population, slope, roads, Military bases and vulnerable locations with the state variable which is a dichotomous variable planned. Accuracy is measured by the area under the ROC curve. The area of sensitivity of 1 represents perfect test. Looking at the Table 26 for area under ROC curve for change in the planned settlement as seen in Table 27 population has a predictive accuracy of greater than 0.5 and slope and military basis have a predictive accuracy of almost 0.5.

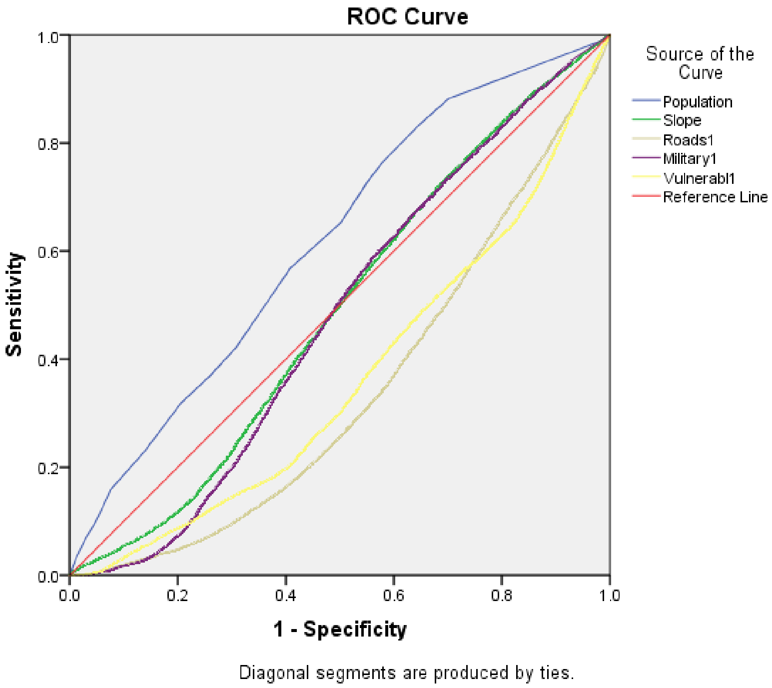


Figure 14 ROC curve for Binary logistic regression on change in planned settlement

Table 27 Area under ROC curve for regression on change in planned settlements

Area Under the Curve	
Test Result Variable(s)	Area
Population Density	0.617
Slope	0.490
Roads1	0.342
Military1	0.477
Vulnerabl1	0.364
The test result variable(s): Population, Slope, Roads1, Military1, Vulnerabl1 has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.	

The graph in Figure 15 represents the ROC test for change in the unplanned settlements. The areas under ROC curve seen in Table 28 shows population density and slope have a predictive accuracy of greater than 0.5.

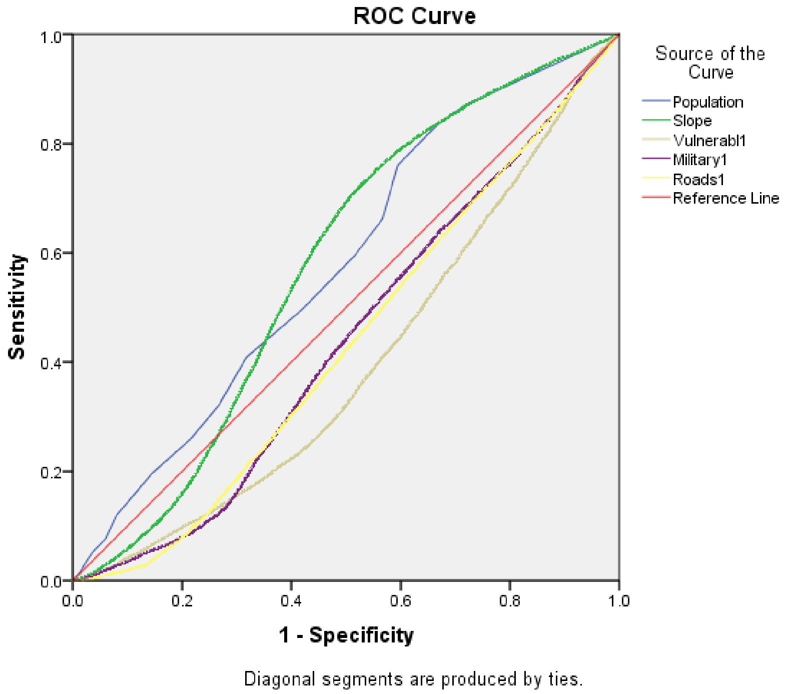


Figure 15 ROC curve for Binary logistic regression on change in unplanned settlement

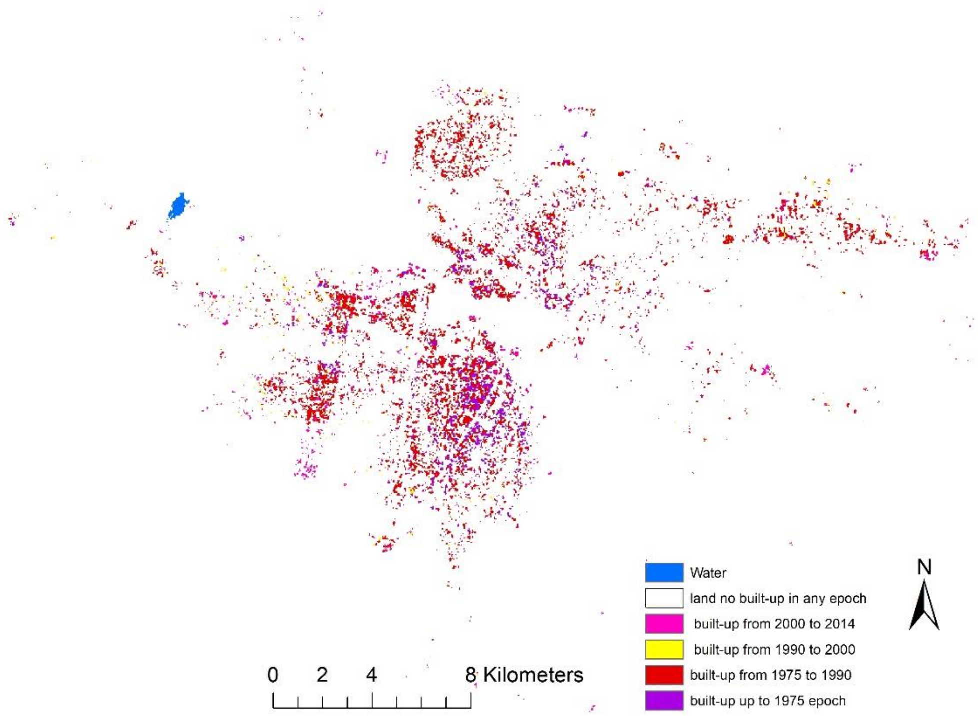
Table 28 Area under ROC curve for regression on change in unplanned settlement

Area Under the Curve	
Test Result Variable(s)	Area
Population Density	0.579
Slope	0.579
Vulnerabl locations	0.396
Military bases	0.440
Roads	0.434
The test result variable(s): Population, Slope, Vulnerabl1, Military1, Roads1 has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.	

4.6. Discussion

The main objective of the research is to study the urban growth of the city of Kabul and the factors influencing this growth. The research results are discussed on the basis of the objectives as given in section 1.2.2.

The first sub objective of mapping urban growth using multi temporal earth observation- based data was fulfilled by using high resolution aerial photograph (0.4 meter) acquired from CRIDA and high resolution Built-up layer (1 meter) extracted from IKONOS imagery. The aerial photograph consisted of several scenes which had to be mosaicked and subset to the extent of the study area. The scenes were mosaicked without color balancing and histogram matching to retain the raw pixel values. Random forest classification was performed using 35 image features which included terrain features, texture features, spectral features and structural features. It was difficult to classify the built-up classes because of the similarities in the different built-up class. The roof tops of the unplanned settlements and the planned settlements are of similar texture and the classification resulted in a lot of misclassified pixels. Standard products like the global human settlement layer are also not good as can be seen from the Fig 16 a lot of built-up areas has not been captured due to the spectral similarities between the built and the non-built. To improve the classification the four built-up classes were combined into 2 classes planned and unplanned. The global human settlement layer for the year 2001 with 1 meter resolution extracted from IKONOS is used to map the change in planned and unplanned from 2001 to 2017. The accuracies of the classified aerial photograph with 3 classes and the Built-up layer extracted from IKONOS is comparable with 70% and 75 % accuracies respectively. This helped in improving the regression model. The urban morphology of the city of Kabul is complex therefore the availability of infra-red band of appropriate resolution would have helped for a more accurate classification.



Source: https://ghsl.jrc.ec.europa.eu/datasets.php
Figure 16 Global Human Settlement Layer

To analyze physical, socioeconomic, accessibility and conflict related factors of urban growth the most important factors which are specific to conflict zone were chosen. The factors Educational institutions, Embassies and Ministries, Hospitals, Mosques and hotel and restaurants are highly correlated and have a Variance inflation factor greater than 10 because of their locational similarity therefore all the factors are important and are merged into one factor as vulnerable locations. As can be seen from the results of the regression Population density, slope and military bases have a high odds ratio for both planned and unplanned settlements. The variables with high odds ratio for change in the planned settlements is shown Table 28.

Table 29 Factors with the highest odds ratio for change in the planned settlements

Variables	B	S.E.	Wald	df	Sig.	Exp(B)
Population Density	0.039	0.003	152.158	1	0.000	1.040
Slope	-0.003	0.001	6.539	1	0.011	0.997
Military bases	-0.180	0.077	5.456	1	0.020	0.835

The variables with high odds ratio for change in the unplanned settlements in shown in the Table 29

Variables	B	S.E.	Wald	df	Sig.	Exp(B)
Population Density	0.026	0.003	91.140	1	0.000	1.026
Slope	0.000	0.000	1.768	1	0.184	1.000
Military bases	-1.028	0.062	276.338	1	0.000	0.358

The overall predicted percentage accuracy for the change in planned settlement is 72% and for the change in unplanned is 65%. The fitness of the model is evaluated with ROC curve and predictive accuracy for change in planned is the maximum for the variable population density and for unplanned it is population density and slope.

5. CONCLUSION AND RECOMMENDATION

Urban growth is a phenomenon happening all over the world but what are its dynamics in a conflict zone like Kabul needs to be studied. To study this phenomenon built-up layer of 1 meter resolution for 2001 i.e. the start of the conflict is provided by JRC and the aerial photograph for 2017 is classified using Random Forest classifier. There are 35 image features used for the classification. The features with high permutation importance and GINI decrease are the terrain features DEM and slope, texture features GLCM second moment, mean, variance and contrast, spectral feature the visible atmospheric resistance index and the original bands 1 and 3 of the aerial photographs. The overall accuracy achieved for the classification with 5 classes is 61%. The amount of growth in the planned settlements is 1.25 times and in the unplanned settlements is 4.5 times from 2001 (start of the conflict) to 2017. The tremendous growth in the unplanned settlements is mainly due to security and economic opportunities in the city. The growth in the unplanned settlement is towards the west and northern parts of the city. On the other hand, the growth in the planned settlements is mainly towards the central and eastern parts of the city. The complexity of the urban form made it difficult to extract the built-up and non-built-up satisfactorily because of the lack of contrast in the classes. It would be helpful to use an infrared image where NDVI could be used to separate vegetation from built-up. The roof of the building for planned and unplanned were similar in Kabul as compare to other cities where they are distinct and can be easily extracted. Kabul is a sprawling city because both planned and unplanned settlements are low rise and low density. The conversion of agricultural lands into built-up also makes the city sprawl.

The factors influencing the growth are chosen keeping in mind the conflict situation in the country. The factors included population density, slope, military basis, road network and the locations that have been attacked in past and have a high probability to be attacked which includes embassies, ministries, educational institutions, hospitals, mosques, hotels and restaurants. The major factors influencing the unplanned growth is population density and slope. The main factors influencing the planned growth is population density and proximity to military bases.

The results of the regression show that the predictive accuracy for the planned settlements is 72.3% and the predictive accuracy of the unplanned settlements is 65%. There is a positive correlation between unplanned growth and slope which indicates that there are unplanned settlements on steep slopes. The expansion of unplanned growth on the west is because of the availability of agricultural land which is being converted into built-up and towards the north east the steeper slopes are converted into unplanned settlements. The central part of the city is planned and has high density of built-up compared to the other parts of the city. The central and eastern parts of the city consist of Ministries, Embassies, hotels, guesthouses for expatriates and government housing built by the Russians for the government employees. The eastern parts of the city consist of international military bases and therefore it is being developed by private builders and the government for provision of housing for the growing population. The model is successful in explaining the growth and the patterns of growth in the city of Kabul which has absorbed by now 3 times the population that existed in 2001.

Recommendation

1. The model can be experimented using other sources of data like high resolution satellite images of different sensors which may be available for the region.
2. The availability of infrared band will be helpful in the classification of urban forms.

3. Random forest algorithm requires high computation time and resources for processing as such they require systems with high configuration to perform classification. The classification can be performed using other algorithms to check for the computation time and processing and the verification of the Random Forest algorithm.
4. To verify the results of the logistic regression other methods such as cellular automata could be performed using the same data.
5. Collaborative program with CRIDA (Capital Region Independent Development Authority) could help them in planning their urban expansion

6. REFERENCES

Afghanistan GDP Annual Growth Rate. (n.d.). Retrieved August 27, 2018, from https://tradingeconomics.com/afghanistan/gdp-growth-annual

Ahmadi, A. S., & Kajita, Y. (2016). Evaluating urban land expansion using geogrpahic information system and remote sensing in Kabul city, Afghanistan. *International Journal of Geomatics and Civil Engineering*, *10*(9), 945–952.

Ahmadi, A. S., & Kajita, Y. (2017). Evaluation of urban land development direcction in kabul city, Afghanistan. *World Academy of Science, Engineering and Technology International Journal of Urban and Civil Engineering*, *11*(2), 152–162.

Aid Effectiveness in Afghanistan. (2018). *Swedish Committe for Afghanistan*. Retrieved from https://reliefweb.int/sites/reliefweb.int/files/resources/aid_effectiveness_in_afhganistan_march_2018_0.pdf

Aithal, B. H., & Ramachandra, T. V. (2016). Visualization of Urban Growth Pattern in Chennai Using Geoinformatics and Spatial Metrics. *Journal of the Indian Society of Remote Sensing*, *44*(4), 617–633. https://doi.org/10.1007/s12524-015-0482-0

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Applying Edge Detection To Feature Extraction And Pixel Integrity. (n.d.). Retrieved April 13, 2019, from https://medium.com/hd-pro/applying-edge-detection-to-feature-extraction-and-pixel-integrity-2d39d9460842

Asian Development Bank. (2017). *Afghanistan: Country Partnership Strategy (2017-2021).* Retrieved from https://www.adb.org/documents/afghanistan-country-partnership-strategy-2017-2021

Atak, B. K., Erdogan, N., Ersoy, E., & Nurlu, E. (2014). ANALYSING THE SPATIAL URBAN GROWTH PATTERN BY USING LOGISTIC REGRESSION IN DIDIM DISTRICT, *1876*(4), 1866–1876.

Atlas of Afghan City Regions. (2016). *Government of Afghanistan*.

Bhandari, A. K., Kumar, A., & Singh, G. K. (2012). Feature Extraction using Normalized Difference Vegetation Index (NDVI): A Case Study of Jabalpur City. *Procedia Technology*, *6*, 612–621. https://doi.org/10.1016/j.protcy.2012.10.074

Bizhan, N. (2018). Aid and state-building, Part II: Afghanistan and Iraq. *Third World Quarterly*, *39*(5), 1014–1031. https://doi.org/10.1080/01436597.2018.1447369

Bürgmann Tatjana. (2015). *Investigating the Separability of Slums in Mumbai based on RADARSAT Images using Object-based Image Analysis Methods*. Retrieved from https://elib.dlr.de/99729/1/Bürgmann_BA.pdf

Cheng, J., & Masser, I. (2003). Urban growth pattern modelling : a case study of Wuhan city , PR China Urban growth pattern modeling : a case study of Wuhan city , PR China. *Landscape and Urban Planning*, *62*(August), 199–217. https://doi.org/10.1016/S0169-2046(02)00150-0

Duque, J. C., Patino, J. E., & Betancourt, A. (2017). Exploring the potential of machine learning for automatic slum identification from VHR imagery. *Remote Sensing*, *9*(9), 1–23. https://doi.org/10.3390/rs9090895

Feng, Y., Liu, Y., & Batty, M. (2016). Modeling urban growth with GIS based cellular automata and least squares SVM rules: a case study in Qingpu–Songjiang area of Shanghai, China. *Stochastic Environmental Research and Risk Assessment*, *30*(5), 1387–1400. https://doi.org/10.1007/s00477-015-1128-z

Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*, *80*(1), 76–87. https://doi.org/10.1016/S0034-4257(01)00289-9

GLCM Texture Feature. (n.d.). Retrieved May 12, 2019, from https://support.echoview.com/WebHelp/Windows_and_Dialog_Boxes/Dialog_Boxes/Variable_properties_dialog_box/Operator_pages/GLCM_Texture_Features.htm

Gov.UK. (n.d.). Terrorism - Afghanistan travel advice - GOV.UK. Retrieved March 14, 2019, from https://www.gov.uk/foreign-travel-advice/afghanistan/terrorism

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*(4), 1164–1176. https://doi.org/10.1109/JSTARS.2012.2190383

Guan, H., Yu, J., Li, J., & Luo, L. (2012). Random Forests-Based Feature Selection for Land-Use Classification Using Lidar Data and Orthoimagery. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XXXIX-B7*(September), 203–208. https://doi.org/10.5194/isprsarchives-xxxix-b7-203-2012

Infrastructure. (2018). Retrieved June 7, 2018, from https://www.usaid.gov/afghanistan/infrastructure

Jochem, W. C., Bird, T. J., & Tatem, A. J. (2018). Identifying residential neighbourhood types from settlement points in a machine learning approach. *Computers, Environment and Urban Systems*, *69*(September 2017), 104–113. https://doi.org/10.1016/j.compenvurbsys.2018.01.004

Kucsicsa, G., & Grigorescu, I. (2018). Urban Growth in the Bucharest Metropolitan Area: Spatial and Temporal Assessment Using Logistic Regression. *Journal of Urban Planning and Development*, *144*(1), 05017013. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000415

Kuffer, M., Pfeffer, K., Sliuzas, R., & Baud, I. (2016). Extraction of Slum Areas From VHR Imagery Using GLCM Variance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(5), 1830–1840. https://doi.org/10.1109/JSTARS.2016.2538563

Mahamud, M. A., Samat, N., & Noor, N. M. (2016). Identifying factors influencing urban spatial growth for the George Town Conurbation. *Planning Malaysia*, *14*(February 2017), 95–106. https://doi.org/10.21837/pmjournal.v14.i5.186

Maini, R., & Dr, &. (n.d.). *Study and Comparison of Various Image Edge Detection Techniques*. *Himanshu Aggarwal International Journal of Image Processing (IJIP)*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.301.927&rep=rep1&type=pdf

Mielke, K., Bicc, \, & Grawert, E. (2016). Why Afghanistan is not a safe country of origin. Retrieved from https://www.bicc.de/uploads/tx_bicctools/policy_brief_1_2016_e_01.pdf

Migration and Urban Development in Kabul: Classification or Accommodation. (2012). Retrieved from http://appro.org.af/preview/migration-and-urban-development-in-kabul/

Mitchell, D. F. (2017). NGO Presence and Activity in Afghanistan, 2000–2014: A Provincial-Level Dataset. *Stability: International Journal of Security and Development*, *6*(1), 5. https://doi.org/10.5334/sta.497

Nong, Y., & Du, Q. (2011). Urban growth pattern modeling using logistic regression. *Geo-Spatial Information Science*, *14*(1), 62–67. https://doi.org/10.1007/s11806-011-0427-x

Pradhan, B. (2017). *Spatial Modeling and Assessment of Urban Form*. (B. Pradhan, Ed.), *Spatial Modeling and Assessment of Urban Form*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-54217-1

Rasmussen, S. E. (2014). Kabul – the fifth fastest growing city in the world – is bursting at the seams | Cities | The Guardian. Retrieved June 7, 2018, from https://www.theguardian.com/cities/2014/dec/11/kabul-afghanistan-fifth-fastest-growing-city-world-rapid-urbanisation

Ruiz Hernandez, I. E., & Shi, W. (2018). A Random Forests classification method for urban land-use mapping integrating spatial metrics and texture analysis. *International Journal of Remote Sensing*, *39*(4), 1175–1198. https://doi.org/10.1080/01431161.2017.1395968

Schug, F., Okujeni, A., Hauer, J., Hostert, P., Nielsen, J., & van der Linden, S. (2018). Mapping patterns of urban development in Ouagadougou, Burkina Faso, using machine learning regression modeling with bi-seasonal Landsat time series. *Remote Sensing of Environment*, *210*(March), 217–228. https://doi.org/10.1016/j.rse.2018.03.022

United Nations. (2018). *The speed of urbanization around the world. DESA, UN*. Retrieved from https://population.un.org/wup/Publications/Files/WUP2018-PopFacts_2018-1.pdf

World Bank. (2018). Afghanistan Overview. Retrieved June 7, 2018, from http://www.worldbank.org/en/country/afghanistan/overview