MAPPING INFORMAL SETTLEMENTS USING MACHINE LEARNING TECHNIQUES, OBJECT-BASED IMAGE ANALYSIS AND LOCAL KNOWLEDGE

Khlood Ghalib Alrasheedi, Ashraf Dewan, Ahmed El-Mowafy

School of Earth and Planetary Sciences (EPS), Spatial Sciences Discipline, Curtin University, Perth 6102, Australia

ABSTRACT

The existence of informal settlements in Riyadh City, the Kingdom of Saudi Arabia (KSA), has given rise to some urban planning issues. To provide improvements to mapping and planning processes, the current study aims to evaluate and characterize informal settlements within the city using object-based machine learning (ML) techniques (specifically, Random Forest (RF) and Support Vector Machine (SVM)), expert knowledge (EK) and satellite data. An examination of four defined locales has produced a comprehensive, local, informal settlement ontology. Four main categories (shape, geometry, texture, and pattern) were used to build the ontological framework. Expert local knowledge was employed to produce a ruleset to accurately identify and map these areas. Specific indicators identified by the specialists were used in a combined object-based ML and image analysis (OBIA) approach, with high-resolution worldview-3 imagery used as input data. Results demonstrated that combining EK and ML with remotely sensed data can efficiently, effectively and accurately distinguish informal settlement areas. This work has shown that an object-based ML technique (RF), in combination with EK about important local environment indicators, provides a useful method for mapping informal settlements.

***Index Terms*—** Local experts, OBIA, random forest, support vector machine, informal settlements

1. INTRODUCTION

The presence of unplanned settlements, colloquially known as informal settlements, can degrade the local environment, limited housing, and high opportunities for residents [1]. These types of settlements have been studied previously in Saudi Arabia, however no mapping of these has been conducted to date in Riyadh [2, 3]. The development of appropriate indicators for identifying unique features of local informal settlements using the knowledge of local experts was not used in any previous research. In previous studies, per-pixel classification techniques have been used to identify informal settlements from high-resolution satellite images. These include Discrete Wavelet Frame Transform (DWFT) [4], Grey Level Co-occurrence Matrix (GLCM) [5], and Local Binary Patterns (LBP) [6].

Object-based image analysis (OBIA) using very high resolution (VHR) imagery is frequently used to map informal settlements [7]. OBIA does, however, have some limitations. Random Forest classifier (RF) and Support Vector Machine (SVM), which both produce reliable classification results, are commonly used with VHR images [5]. Combining both ML and OBIA enhances mapping of informal settlements and maximizes the benefits of each of the respective techniques [5].

The Generic Slum Ontology (GSO) is an ontological framework developed by Kohli, Sliuzas [8] as an informal settlement structural concept that involves the use of remote sensing data. The utilization of this concept in Riyadh city involved characterization of the following indicators – geographical location, neighborhood, shape, density, building and access networks.

A substantial amount of research, in a wide variety of geographical areas, has been undertaken on this topic, however there is still no simple, automated method available for informal settlements mapping, particularly using VHR imagery. The current work involves combining ML and OBIA to take the advantage of the specific features of each method (Han et al., 2020; Kuffer, Pfeffer, & Sliuzas, 2016).

The aim of this study is to explore and construct an ontology of informal settlements in Riyadh using object-based methods, local expert knowledge, and remote sensed data. Object-based ML and OBIA approaches will be used to refine the informal settlement ontology using suitable indicators identified by local specialists.

2. MATERIALS AND METHODS

2.1 Description of the Case Areas

The study area is located in Riyadh, the capital and largest city in (KSA). It is situated approximately 600 meters above mean sea level in the Najd Plateau. Four neighborhoods were chosen as cases in this work. These are: Al Shomaisi (1.49 km^2), Meekal (0.21 km^2) Al Dirah (0.24 km^2) and Al Dubiya (1.56 km^2).

2.2 Dataset

The input data included multispectral worldview-3 imagery, road networks and informal settlement boundary. A digital elevation model (DEM) and digital surface model (DSM) were also used, along with an EK survey to determine indicators for the determination of informal settlement.

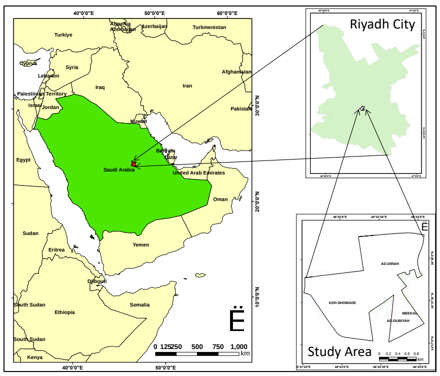

Figure 1. Case studies location.

2.3. Ontological framework

The GSO concept was used as a template in the development of a local ontology of informal settlements (LOIS) for Riyadh. Sixteen indicators were identified to use with processing of satellite imagery. The LOIS framework defined three levels of interest: object, settlement, and environs. Subject experts with knowledge of local conditions were interviewed to accurately and effectively identify suitable indicators for use in the identification process of informal settlements. The final classification results provided good evidence of the suitability of this approach. Confirmation of classification accuracy was used as justification for further work, combining OBIA and ML techniques.

Commented [A1]: Where is your evidence?

2.4. Object-based RF and SVM classifications

Segmentation was accomplished using multi-resolution techniques and the specific identification of spectral differences. A scale parameter (SP) value of 30 was chosen and weighting values of 0.3 and 0.5 were given to shape and spectral reflectance compactness, respectively [8]. Six types of land use and cover were identified and classified. These consisted of road networks, vegetation, shadows, vacant areas, informal and formal settlements. The roofs of buildings were extracted using DSM and GLCM entropy data. To extract the lacunarity of housing structures and to improve the textural visibility of built-up regions, GLCM homogeneity, GLCM contrast, GLCM correlation, and GLCM mean were also utilized. Shadows and vacant areas were identified using visible brightness features. Vegetation was mapped using normalized difference vegetation index (NDVI).

Approximately 6000 points were randomly generated within each class. Each point was converted to a polygon with a radius of 0.4 m and combined into a set, while 1800 polygons were set aside and used for training. Training data and class names were used as variable inputs for prediction purpose. Sixteen indicator features were also used in the explanatory training. 500 trees (maximum tree depth 30) were used for the RF classification. The importance of each indicator used in the RF classification was assessed using an out of box (OOB process. 70% of the data was used for prediction and 30% for training [9]. SVM was used in the settlement differentiation process. For classification, a C-SVM with a value of 1 was employed. The classification used a radial basis function (RBF) kernel, with width inversely proportional to its variance [5]. The overall accuracy assessment of the study was performed using a confusion matrix.

3. RESULTS AND DISCUSSION

Satisfactory segmentation and classification results were achieved using OBIA using the defined informal settlement indicators and LOIS indices (Figure 2). OBIA segmentation and object-based RF and SVM were also implemented (Figures 3 and 4).

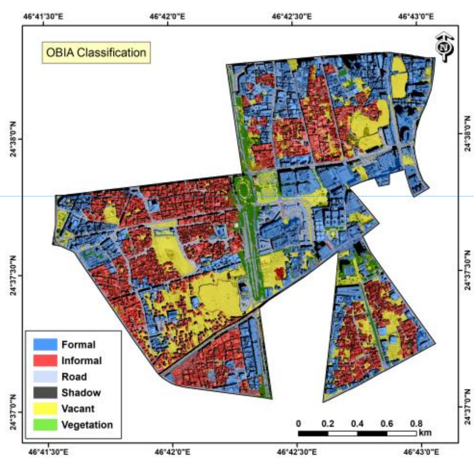

Figure 2. Classified image of informal and formal settlements in Riyadh city, based on OBIA.

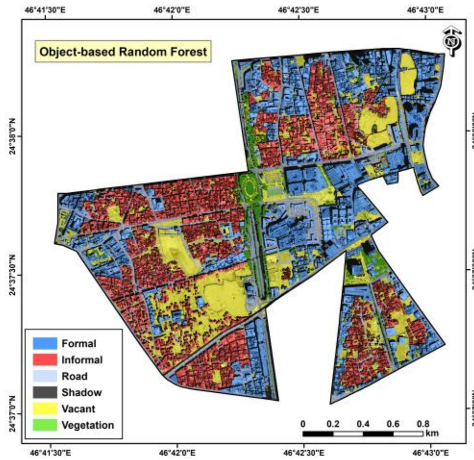

Figure 3. Object-based RF classified image of informal and formal settlements.

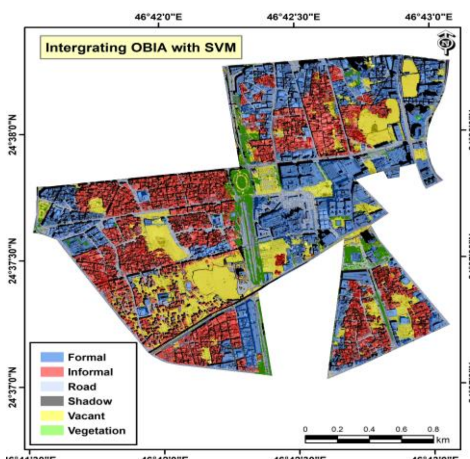

Figure 4. Object-based SVM classified image of informal and formal settlements.

Border index was ranked as the first important variable and vegetation was ranked as the second one using RF and as the third most important using SVM (Figure 5). This significantly affected accuracy of the classification process. NDVI appears to be a good approach to use in identifying vegetation from other indicators using red and infrared bands. Built-up area indicators were ranked among the third most important indicator in the variable importance graph generated by RF and fourth in SVM. This result agrees with studies conducted by [10, 11] but not with [5] who ranked this variable amongst the lowest important indicator. Note that the previous research focused on slums.

The border index appeared to be the most important indicator in the RF classification. This feature determined the boundary of buildings in the VHR images during the segmentation work. It was noted, however, that in SVM the border index was ranked as the eighth most important indicator. This can be referred to as a Radial Basis Function (RBF) approach when applied to classification using the SVM.

This study found that the GLCM approach was used to map the five texture techniques used in segmentation to create the texture of each object. In regards classifying VHR imagery, GLCM homogeneity and GLCM entropy appear to be the most important indicators for identifying density and textural components. [12] GLCM homogeneity ranks first among the GLCM textural features due to its ability to capture the spatial distribution characteristics of different built-up areas, and its ability to discriminate between the homogeneous and heterogeneous nature of these areas.

GLCM entropy appeared to be the best indicator for extracting roof of large buildings. This has been used in some studies to map informal settlement layouts and other land cover classes [13]. At the settlement level, GLCM mean, GLCM contrast and GLCM correlation were the approaches that least impacted the final classification.

The dwellings shape indicator was ranked the second most important indicator in SVM. For this reason, SVM may be useful in determining the shape and area of built-up zones (see Figure 5). The standard deviation of the blue band indicator was used to measure data dispersal in relation to mean brightness in this band. It was ranked highly in both the RF and SVM process. The blue and red bands appear to be the best VHR bands for mapping informal settlements.

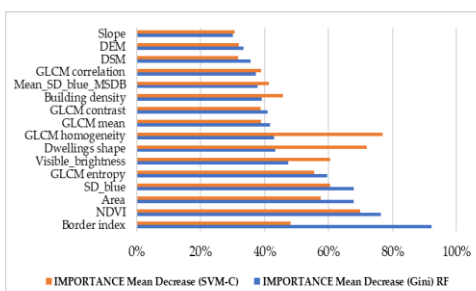

Figure 5. Importance of informal settlements indicators in RF and C-SVM classification techniques.

The accuracy of training areas for RF and SVM were calculated for all classes and the results are displayed in Figure 6. Overall accuracy of the study was regarded as acceptable (98% with a corresponding kappa value of 97% for RF, and 95% and kappa value of 91% for SVM).

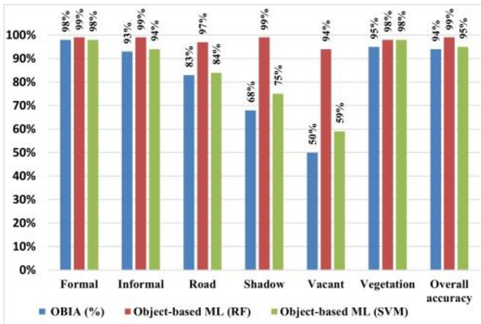

Figure 6. Overall accuracy of Object-based RF and SVM classification.

4. CONCLUSION

The purpose of this study was to use OBIA-based RF, SVM, remotely sensed data, and expert knowledge to develop an ontology of informal settlements for Riyadh. Classifications developed using RF and SVM techniques were used to produce a map with six explanatory classes, namely formal and informal settlements, road networks, vacant areas, areas of shadow and vegetated areas. The study has shown that object-based RF combined with expert knowledge regarding the use of suitable indicators can prove useful in mapping informal settlements within the Middle East cities.

.

5. REFERENCES

1. Khraif, R., et al., Residential Satisfaction in Shantytowns of Riyadh City, Saudi Arabia: Levels and Determinants. International Journal, 2018. 5(3).
2. Breengy, A. and N.A. Yusof, Building-Related Health Issues in an Unsustainable Neighbourhood–a Study of a Slum Area in Jeddah, Saudi Arabia. The Arab World Geographer, 2018. 21(2-3): p. 141-153.
3. Fallatah, A., et al., Mapping informal settlement indicators using object-oriented analysis in the Middle East. International journal of digital earth, 2019. 12(7): p. 802-824.
4. Prabhu, R., B. Parvathavarthini, and R. Alagu Raja, Slum extraction from high resolution satellite data using mathematical morphology-based approach. International Journal of Remote Sensing, 2021. 42(1): p. 172-190.
5. Leonita, G., et al., Machine learning-based slum mapping in support of slum upgrading programs: The case of Bandung City, Indonesia. Remote sensing, 2018. 10(10): p. 1522.
6. Matarira, D., O. Mutanga, and M. Naidu, Texture analysis approaches in modelling informal settlements: A review. Geocarto International, 2022 (just-accepted): p. 1-32.
7. Mahabir, R., et al., A critical review of high and very high-resolution remote sensing approaches for detecting and mapping slums: Trends, challenges and emerging opportunities. Urban Science, 2018. 2(1): p. 8.
8. Kohli, D., et al., An ontology of slums for image-based classification. Computers, Environment and Urban Systems, 2012. 36(2): p. 154-163.
9. Feng, Q., J. Liu, and J. Gong, UAV remote sensing for urban vegetation mapping using random forest and texture analysis. Remote sensing, 2015. 7(1): p. 1074-1094.
10. Pratomo, J., et al., Coupling uncertainties with accuracy assessment in object-based slum detections, case study: Jakarta, Indonesia. Remote sensing, 2017. 9(11): p. 1164.
11. Kohli, D., et al., Identifying and classifying slum areas using remote sensing. University of Twente Faculty of Geo-Information and Earth Observation (ITC): Enschede, The Netherlands, 2015.
12. Kuffer, M., et al., Extraction of slum areas from VHR imagery using GLCM variance. IEEE Journal of selected topics in applied earth observations and remote sensing, 2016. 9(5): p. 1830-1840.
13. Duque, J.C., J.E. Patino, and A. Betancourt, Exploring the potential of machine learning for automatic slum identification from VHR imagery. Remote Sensing, 2017. 9(9): p. 895.