Land Suitability Analysis to Assess the Potential of Public Open Spaces for Urban Agriculture Activities

Dissertation submitted in partial fulfillment of the requirements

for the Degree of Master of Science in Geospatial Technologies

February 24, 2020

Maicol Fernando Camargo Hernández

mfcamargoh@gmail.com

https://github.com/mfcamargoh

Supervised by:

Hanna Meyer

Institute for Geoinformatics

University of Münster

Co-supervised by:

Judith Verstegen

Institute for Geoinformatics

University of Münster

and

Pedro Cabral

Information Management School

Universidade Nova de Lisboa

Declaration of Academic Integrity

I hereby confirm that this thesis on *Land Suitability Analysis to Assess the Potential of Public Open Spaces for Urban Agriculture Activities* is solely my own work and that I have used no sources or aids other than the ones stated. All passages in my thesis for which other sources, including electronic media, have been used, be it direct quotes or content references, have been acknowledged as such and the sources cited.

February 24, 2020

I agree to have my thesis checked in order to rule out potential similarities with other works and to have my thesis stored in a database for this purpose.

February 24, 2020

Acknowledgments

The idea about this research came to my mind when I was visiting the urban orchard of my aunt Elena Villamil located in downtown Bogota. I was fascinated by the idea of harvesting our food and the social, environmental and cultural benefits involved in urban agriculture practices. Since that moment, I wanted to develop my thesis related to urban agriculture to benefit the population of my city. For this and much more, infinite thanks to my aunt for inspired me with her passion for this humble and rewarding activity. From my heart, I hope her wisdom and dedication in that small green spot in the middle of that concrete jungle will be maintained for many more years and hopefully will continue spreading around the country. This thesis would have not been possible without the help and support of an important group of people.

First, I would like to thank my supervisor, Professor Hanna Meyer, for her patience and continuous guidance throughout the development of this research. Next, I want to acknowledge my co-supervisor, Professor Judith Verstegen for her constructive commentaries and advice about my work. I also want to express my gratitude to the Erasmus Mundus Master program Science in Geospatial technologies, headed by the professors Marco Painho, Christoph Brox, Joaquin Huerta, and Michael Gould. I am very honored for having been one of the recipients of this Erasmus scholarship.

My sincere thanks to Edgar Lara and the Botanic Garden of Bogotá for the information provided on urban agriculture in the city, and for always been in contact when I needed supervision and advice. Special thanks also to Daniel Vega, for sharing his wisdom about urban agriculture and encourage this research. I also want to thank Diana and Alejandro, for their patience and valuable feedback in different steps of this study. Additionally, I would like to thank my friends Carlos and Violeta for their support, help, motivation, and joy shared during this Master experience. And, my good German friends Moritz and Neele, for extending your solidarity and support when I need it the most in Münster.

Finally, my heartfelt thanks to my family and friends in Colombia, for their love and constantly sent me their affection from abroad. And, my geospatial classmates for all the good moments that we shared in Spain, Portugal, and Germany, this journey wouldn't have been the same without you.

Abstract

In a world increasingly dominated by cities and an accelerated urban sprawl, urban agriculture emerges as an alternative for the continuous stock and food supply that urban population demands. This thesis aimed to identify and evaluate potential available areas in public locations for implementing urban agriculture practices within the urban perimeter of the city of Bogota in Colombia. The methodology was conducted using variables reflecting the physical, environmental and socioeconomic components of the area. Two approaches were implemented to evaluate a land suitability analysis for urban agriculture to alleviate urban poverty by increasing food security and nutrition in the study area. The first approach was based on expert knowledge combining GIS with multicriteria decision making analysis (MCDM) using analytical hierarchical process (AHP) method, estimating that 21% of the study area presents highly suitability conditions for implementing urban agriculture activities. The second approach was developed using supervised machine learning algorithms for classification models based on historical data of the current sites, where urban agriculture activities were being implemented in the city, showing that 18% of the study area is in high suitability conditions for the implementation of urban agriculture activities. Both approaches indicated that the areas of excellent suitability are located in the South and Southwestern parts of the study area, emphasizing its congruence with the areas with the lowest socioeconomic levels in the city.

It was found that approximately 2% of the study area has available spaces in public locations with a significant potential for urban agriculture practices. Three projected scenarios were simulated where 10%, 30% and in the most utopic case 50% of these spaces would be used for urban agriculture activities and the vegetable productivity in tons of five of the most popular crops grown was estimated.

Key Words

Analytic Hierarchy process

K-Nearest Neighbors

Machine Learning

Multicriteria Decision Analysis

Random Forest

Sensitivity Analysis

Support Vector Machine

Urban Agriculture

Acronyms

AHP	Analytic hierarchy process
KNN	K-nearest neighbors
MCDM	Multicriteria Decision Making Analysis
ML	Machine Learning
MPI	Multidimensional Poverty Index
MPV	Monetary Poverty
OAT	One at a time
POS	Public Open Spaces
PSI	Public Space Indicator
RALSE	Residential Areas with Low Socioeconomic Level
RF	Random Forest
SVM	Support vector machine
UA	Urban agriculture
UMNRATE	Unemployment Rate

Table of Contents

Index of Tables

Index of Figures

1. INTRODUCTION

The migration from the countryside to the city is a phenomenon that has affected Colombia for several decades, especially for political reasons related to forced displacement and violence (McEniry, Samper-Ternent, & Cano-Gutierrez, 2019). Forced displacement has been one of the most serious consequences of the Colombian armed conflict, situating it as the second country in the world with the highest number of internally displaced persons, exceeded only by Syria (Toole, 2019). Additionally, due to the current political instability and the economic crisis that is facing Venezuela, nearly 1.5 million Venezuelans citizens and refugees had emigrated to Colombia (Pantoulas & McCoy, 2019). Cities provide economic, social and cultural opportunities that have always attracted migrants in search of a better quality of life and opportunities (Goldscheider, 2019). This high rate of migration is usually accompanied by a phenomenon defined as the "urbanization of poverty" which leads to a displacement of poverty from rural to urban areas (Ingersoll, 2012)

Bogotá as the main capital and focus of the most important administrative, political, economic and industrial activities is an ideal candidate for refugees and search for new opportunities. Therefore, as a result of the high number of migrants and the constant urban expansion, the urban area of Bogotá represents a potential candidate for the development of sustainable technologies and methods that helps to mitigate the high poverty rates, contributing with the development of local economies as an alternative of sustenance for unemployed people (Orsini et al., 2013). It is in this aspect where Urban Agriculture (UA) emerges as an alternative to carry out productive activities that respond to the basic needs of the communities, contributing to mitigating problems related to poverty alleviation and environmental degradation (Badami & Ramankutty, 2015; Lee-Smith, 2010). Multiple benefits have been associated with the development and integration of UA in the cities, including the use of clean and environmentally friendly energies, continue access to healthy food, reduction in air pollution and soil erosion, and improvement food and nutrition security by the increase of the food supply for the urban population (Foster & Rosenzweig, 2007).

Most cities have available areas and underutilized spaces that might be used for UA (Thomaier et al., 2015). In the inner cities, urban farmers still make use of schools, churches, unused land and road/rail sides but they have all come with various challenges, especially when land is scarcely available for agricultural purposes in cities such as Bogotá (Van Veenhuizen & Danso, 2007). Therefore, this thesis considers the implementation of a land suitability analysis within the urban perimeter of Bogotá, to identify potential available areas in public spaces for the development of UA activities by the comparison of an approach based on a subjective method using a Multicriteria Decision Making Analysis (MCDM) and a approach based on an objective assessment derived from historical data using machine learning techniques.

1.1 RESEARCH QUESTIONS

This thesis formulated the following research questions:

- Where are the most suitable areas for developing urban agriculture practices located in the city of Bogotá?
- What are the most relevant criteria for urban agriculture derived from expert knowledge and data-driven approach and how do the approaches compare?
- What could be the possible vegetable production of public open spaces within the study area?

1.2 AIM

To answer the research questions defined the main objective of this thesis is:

- Implement a Land Suitability Analysis to identify potential available areas for urban agriculture practices in public open spaces using an expert knowledge approach based on Multicriteria Decision Making Analysis (MCDM) and a data-driven approach based on machine learning methods

1.3 OBJECTIVES

To achieve the formulated research questions, this thesis proposed the following Specific objectives:

- Review and evaluate the potential of machine learning methods for Land Suitability Analysis
- Design and implement a Land Suitability Analysis to identify potentially suitable areas for UA activities using an expert knowledge approach based on MCDM and a data-driven approach using machine learning techniques
- Assess the performance of the methods implemented in both approaches based on existent data
- Compare the level of correspondence from the proposed approaches

2. LITERATURE REVIEW

2.1 RELATED WORKS

There are several kinds of research, city councils, and countries governments that have tried to encourage and incentivize interest in developing or improving UA in their territories. Dongus and Drescher (2009) produced a map of vegetable production on open spaces in Dar es Salaam, Tanzania using aerial photography and Global Positioning Systems (GPS) instruments, showing the spatial changes from 1992 to 1999. An inventory of all open spaces with information related to their location, size and other fieldwork attributes were integrated into a GIS database. The use of remote sensing techniques on satellite images of high. Ermini et al. (2017) developed a methodology to generate information on urban and suburban agricultural activities in the metropolitan area of Santa Rosa-Toay, Argentina. Urban agriculture was included through a participation system based on interviews and cross-validating information using google maps. The development of a combined methodology (quantitative and qualitative) allowed collecting relevant information and points of view considered by local farmers. Uy and Nakagoshi (2008) implemented a land suitability analysis to quantified suitable sites for developing urban green spaces in Hanoi, Vietnam. A GIS-based multicriteria decision making analysis (MCDM) using the analytic hierarchy process (AHP) and an ecological factor threshold method were combined to create a composite map represented as a suitable green map. This was then compared with the 2020 Hanoi Master Plan showing a high grade of compatibility. McClintock et al. (2013) assessed the potential contribution of vacant land to urban vegetable production and consumption in Oakland, California. The contribution of vegetable production for four different land-use scenarios was estimated using census data and a vacant land inventory (including vacant lots, open space, and underutilized parks) with agricultural potential were identified using GIS and aerial imagery of the city. The main purpose of this study was to identify vacant parcels (public and private) that might represent potential sites for food production. According to this, 486.4 ha of public land and 136.4 ha of private land of 756 individual tax parcels were registered as potential candidates for use for vegetable production.

La Rosa and Privitera (2014) developed a methodology for sustainable planning of new forms of agriculture in urban contexts in the municipality of Catania, Italy using GIS-MCDM model and relative spatial indicators. The study validated the suitability of land-use transitions of current Non-Urbanized Areas to New Forms of UA, introducing scenarios to increase food production and access to green spaces. Another example of suitability analysis for UA was implemented by Hemakumara (2015) in Colombo, Sri Lanka. In this study, several indicators to measure UA suitability were identified based on a GIS combined with MCDM using the AHP method. This process allowed the development of suitable decision scenarios for different UA practices in the study area.

Recent studies have involved more innovative methods for land suitability analysis such as machine learning techniques. Heumann et al. (2011) adapted a niche theory to a human-managed landscape in a land suitability modeling using the Maximum Entropy model in the Nang Rong District, Northeastern Thailand. Based on a socio-environmental niche where the likelihood of crop occurrence is a function of natural, built, and social environmental conditions that might influence in the determining land use choices in a human-managed landscape, crop occurrences were modeled showing that natural environment is often the dominant factor in crop likelihood, the likelihood is also influenced by household characteristics, such as household assets and conditions of the neighborhood or built environment. Sarmadian et al. (2014) evaluated the potential use of the Support Vector Machines algorithm for land suitability analysis for rainfed wheat in the northwestern province of Qazvin in Iran. The results showed that the most important limiting factors for rainfed wheat cultivation are climatic and topographic conditions. Test data points were used to predict land suitability indices to assess the algorithm performance. The Root Mean Square Error (RMSE) and coefficient of determination (R2) were used as evaluation criteria between the measured and predicted land suitability indices, obtaining values of 3.72 and 0.84 respectively, concluding that Support Vector Machines approach could be a suitable alternative to performance of land suitability scenarios. Mokarram al. (2015) implemented machine learning algorithms for land suitability classification in the northern of Khuzestan province, southwest of Iran. The study investigated the potential of the RotBoost method to land suitability classification and comparison with other methods such as Bagging, Rotation Forest and Boosting techniques to find the best method for land suitability classification, obtaining that RotBoost algorithm was more accurate than the other method and concluding that the use of machine learning methods have a positive implementation in land suitability classification, especially multiple classifier system methods.

Senagi et al. (2017) compared the performance of Parallel Random Forest, Support Vector, Linear Regression, Linear Discriminant Analysis, K Nearest Neighbor, and Gaussian Naïve Bayesian machine learning algorithms for predicting land suitability for sorghum production based on soil properties information in Kenya. Results showed that parallel random forest had better accuracy (0.90) and a lower standard deviation (0.13). The main conclusion was that parallel random forest can optimize the prediction of land suitability for crop production based on soil information.

From the above literature, it can be seen that diverse studies have implemented GIS with multicriteria decision making analysis (MCDM) for land suitability oriented to agriculture or crop production, but few studies have dealt with the use of machine learning techniques in land evaluation, especially in urban agriculture (Sarmadian et al., 2014). Moreover, no studies were found that attempted to evaluate or compare the results of a land suitability assessment from subjective human-based methods such as MCDM against methods based on objective assessments derived from data learning such as machine learning techniques.

2.2 LAND SUITABILITY ANALYSIS

Land suitability refers to the ability of a portion of land to support the production of crops sustainably, advising to grow or not grow a particular crop (Singha & Swain, 2016). Land suitability analysis estimates the suitability of an area for a specific use for each land mapping unit (Senagi et al., 2017). Improve food security and malnutrition can be achieved by encouraging sustainable urban agriculture (Badami & Ramankutty, 2015). Identify suitable and available lands for this particular use, is a complex process that requires multiple decisions relate to the analysis and interpretation of a wide number of variables and criteria (qualitative and/or quantitative with differing importance) from multiple sources of information (Jafari & Zaredar, 2010; Mendas & Delali, 2012). Therefore, land suitability analysis for urban agriculture could be considered as a process of multicriteria decision support (Prakash TN, 2003).

2.2.1 Multicriteria Decision Making Analysis (MCDM)

Multicriteria decision making (MCDM) has been widely applied for decision-makers that have to deal with complex choices for problems that require a selection of the best alternative, according to their preferences from multiple potential candidates (Pavan, 2009). The problem is divided into smaller parts where each one is analyzed separately making it easier for the decision-makers to understand and have confidence about making a decision that involves again of all parts (Malczewski, 2006). Due to the capabilities of spatial data manipulation, extraction, and analysis provided by GIS and the potential for structuring and evaluating decision problems prioritizing alternative decisions in MCDM, both techniques have been combined in a spatial multicriteria decision making in several land suitability analysis (Aldababseh et al., 2018; Liu et al., 2007; Montgomery & Schmidt, 2015; Singha & Swain, 2016). This can view at a basic level as a process that transforms and combines value judgments (coming from the decision-makers preferences) with spatial data to provide information for decision making. (Malczewski, 2006). In this scenario, the main goal is to provide solutions for spatial decision problems with multiple criteria, where each criterion is a spatial data transform into a decision. The problem is decomposed in a hierarchical structure providing a general view of the complex relations in the analysis and provide to the decision-makers to distinguish the level of importance of the criteria (Prakash TN, 2003).

2.2.2 Analytic Hierarchy Process (AHP)

The Analytic Hierarchy Process (AHP) is a multicriteria decision-making approach based on human judgment ability to structure a multicriteria problem as a hierarchical model formed by objectives or main goal, criteria, sub-criteria or variables and alternatives (Setiawan, Sediyono, & A. L. Moekoe, 2014, Saaty, 1977). Saaty(1987) defined that the AHP method has three principles: decomposition, comparative judgments, and synthesis of priorities that can be explained in the following steps.

1. Defining the model structure: decomposition means organizing the problem on different levels. A hierarchical structure is built as a decomposition structure that includes the decision goal(s), main criteria, sub-criteria and alternatives to be used to define land suitability levels (Aldababseh et al., 2018). Figure 1. Example of a

hierarchy of criteria An example of decomposition in a hierarchical structure is shown in Figure 1 (Vargas, 2010).

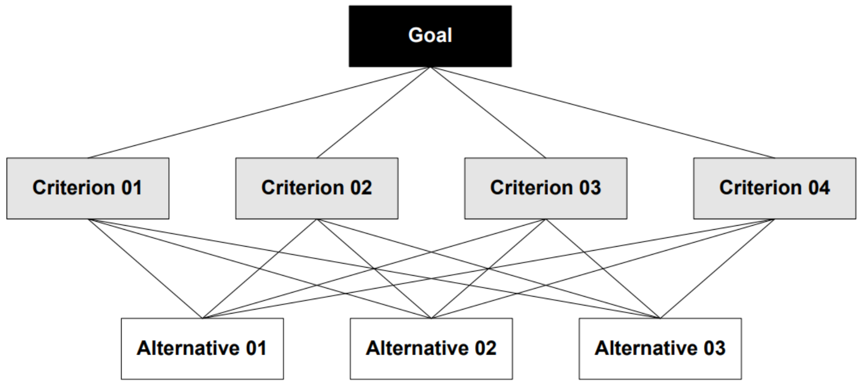

Figure 1. Example of a hierarchy of criteria

2. Standardization of criteria: as was mentioned in section 2.2, a land suitability analysis for urban agriculture dealt with heterogeneous criteria (quantitive/qualitative) that come in different measurement scales. To perform comparative judgments based on expert preferences, it is necessary to convert all criteria into a common domain of measurement. To accomplish this, criteria should be standardized considering the goal and alternatives under evaluation (Prakash TN, 2003).

3. Assigning weights: defining the criterion weights is a fundamental requirement for applying the MCDM/AHP method (Feizizadeh & Blaschke, 2013). A comparison between each criterion under evaluation is carried out based on expert knowledge and literature review to provide the best judgment of their relative importance using a pairwise comparison matrix (Aldababseh et al., 2018). This can be mathematically expressed in the following equation:

$$A = [a_{ij}], i, j = 1, 2, 3 \dots, n; \quad (1)$$

where A is the matrix with a_{ij} elements, in which all elements are compared with themselves, i and j are the criteria with a reciprocity property of $a_{ij} = 1/a_{ij}$ for all i and j.

Intensity of importance	Definition Explanation	Definition Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Weak importance of one over another	Experience and judgment slightly favor one activity over another
5	Essential or strong	Experience and judgment strongly favor one activity over another
7	Demonstrated importance	An activity is strongly favored, and its dominance demonstrated in practice
9	Absolute importance	The evidence favoring one activity over another is of the highest possible order of affirmation
2, 4, 6, 8	Intermediate values between the two adjacent judgments	2, 4, 6, 8 Intermediate values between the two adjacent judgments
Reciprocals	If activity i has one of the above numbers assigned to it when compared with activity j, then j has the reciprocal value when compared with i	

Table 1. Scale for pairwise comparisons (T. L. Saaty, 1977)

The level of importance between all criteria is evaluated using Saaty's s weighting scale shown in Table 1. Weights are estimated by normalizing the pairwise comparison matrix which is obtained by dividing the column elements of the matrix by the sum of each column (Equation 2). Row elements in the obtained matrix are summed, and the total value is divided by the number of elements in the row as is shown in Equation 3:

$$A' = \left[a'_{ij}\right], i, j = 1, 2, 3 \ldots, n \tag{2}$$

where A' is the normalized matrix of A and the a'_{ij} is defined as:

$$a'_{ij} = \frac{a'_{ij}}{\sum_{i=1}^{n} a_{ij}} \tag{3}$$

for all i,j = 1, 2, 3, ..., n. Then, criteria weights are estimated as a priority vector or weight vector (Akinci, Özalp, & Turgut, 2013).

$$w_i = \frac{\sum_{j=1}^{n} a'_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n} a'_{ij}} \tag{4}$$

Weights values are within 0 a 1, and their sum is equal to 1

$$\sum_{i=1}^{n} w_i = 1 \tag{5}$$

4. Consistency: A Consistency Ratio (CR) is used to measure the inconsistency obtained as a result of the expert judgments based on the estimation of the

Consistency Index (CI) that validated the consistency in the pairwise comparison matrix (Aldababseh et al., 2018). CI can be estimated and written as:

$$CI = (\lambda_{max} - n)/(n - 1) \quad (6)$$

where n is the number of elements being compared in the matrix and λ_{max} is the largest eigenvalue of the matrix. Then, CR is calculated as follows:

$$CR = CI/RI \quad (7)$$

where RI is the Random index obtained randomly through experiments using samples for different numbers of elements or criteria (Chivasa, Mutanga, & Biradar, 2019). Table 2 shows the RI for the first 10 samples. To be accepted the CR must be < 10%, otherwise, judgments are considered inconsistent and the expert o decision makers should re-evaluate the pairwise comparison to identify the possible inconsistency and repeat the process until the CR could be acceptable. If the CR is below 10%, judgments are considered consistent (T. L. Saaty, 1977).

N	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
RI	0	0	0.52	0.89	1.11	1.25	1.35	1.40	1.45	1.49

Table 2. The Random Indices (T. L. Saaty, 1977).

5. Model Synthesis: suitability scores defined for the sub-criteria or variables (defined in the standardization step) within each criterion are multiplied with the weights assigned for each criterion to calculate the suitability index and generate the final suitability map (Aldababseh et al., 2018).

6. Making a final decision: stakeholders or the decision-makers should come to a final decision with consideration of all important criteria and the results obtained using the AHP method (Fadhil & Moeckel, 2018).

Among many developed multicriteria decision-making methods, this thesis utilized the AHP method because of its capacity to integrate a large amount of heterogeneous data and simplicity to include different opinions (Y. Chen, Yu, & Khan, 2010; Feizizadeh & Blaschke, 2013). Besides, AHP also is widely known among researchers due to its effective mathematical properties and has been implemented in different studies related to agriculture land-use suitability analysis (Akinci, Özalp, & Turgut, 2013; Montgomery & Schmidt, 2015; Puntsag, Kristjánsdóttir, & Ingólfsdóttir, 2014; Setiawan et al., 2014)

2.3 SENSITIVITY ANALISIS

Decision making is a subjectivity process that is accompanied by uncertainty (Prakash TN, 2003). In the MCDM-AHP method, uncertainty may come from many different sources, such as original data, data processing, criteria selection and judgments of the experts or decision-makers. Experts may not be completely aware of their preferences concerning the criteria or a definition of a unique set of values for the weights is not possible, due to several judgments and preferences coming from multiple experts. The weight estimated for each criterion is one of the most sensitive parameters and a potential contributor to uncertainty in an MCDM - AHP implementation (E. Xu & Zhang, 2013). Therefore, a sensitivity analysis of the weights of input criteria is crucial to reducing uncertainty and increasing the stability of the outputs (Yun Chen, Yu, & Khan, 2013). Sensitivity analysis studies how the variations in input parameters modify the model output (Montgomery & Schmidt, 2015). It can be applied to evaluate how uncertainty in model inputs, influences uncertainty in model predictions. It is considered a good modeling practice to perform validation and calibration of numerical models using sensitivity analysis, to prove the robustness of the final result against small variations in the input data (Crosetto, Tarantola, & Saltelli, 2000).

For the development of this thesis, a combine sensitivity analysis using the One At a Time method (OAT) and GIS techniques was addressed, due to its simple implementation with a low computational cost (Y. Chen et al., 2010). Additionally, the OAT is considered the most straightforward method to validate uncertainty in models, estimating the effect on the evaluation results based on variations in a single input parameter, while holding all other parameters fixed at their nominal values (E. Xu & Zhang, 2013).

2.4 CLASSIFICATION ALGORITHMS

Machine learning techniques are approaches composed of statistical models and algorithms, whose main aim is to learn from the analysis of data (training data) identifying existing patterns and converting this experience into knowledge or expertise to perform a task (Shalev-Shwartz & Ben-David, 2013). There are different categories of machine learning algorithms but usually are classified depending on the learning type (supervised/unsupervised) or learning models (classification, regression, clustering, and dimensionality reduction) or the learning models employed to implement a selected task (Liakos et al., 2018). Random Forest (RF), Support Vector Machine (SVM) and K Nearest Neighbors (KNN) algorithms were selected as they have been implemented as supervised classification models in agriculture land suitability analysis (Heumann et al., 2011; Sarmadian et al., 2014; Senagi et al., 2017). Classification algorithms belong to the category of supervised learning and they are characterized by the use of datasets with labels that generate a predicted class of type discrete (Kuhn & Johnson, 2013)

2.4.1 Random Forest

Random Forest is an ensemble algorithm based on the implementation of multiple decision trees to make classifications and prediction classes (Breiman, 2001). Each tree contains random samples of training data points (from the original data) and each node contains a random subset of predicting variables (features). Additionally, each tree in the forest vote for the classification of a new sample and the final prediction of the algorithm is obtained by the average of votes over the predictions of the individual trees (Kuhn & Johnson, 2013). Due to the construction of ensemble trees, random forest contributes to control variance and overfitting improving the performance of the final prediction (Breiman, 2001). Moreover, it is easy to implement because only two parameters are required: the number of trees (*ntree*) and the number of predicting variables randomly used (*mtry*) at each split (Shalev-Shwartz & Ben-David, 2013). The most common way to tune the performance of RF is by increasing the number of decision trees that the algorithm generates to obtain a more reliable result. However, as the final model consists of a group of decision trees, could be difficult to interpret (Castelli, Vanneschi, & Largo, 2019). Figure 2 (Abilash, 2018) shows an example of classification by RF using four trees.

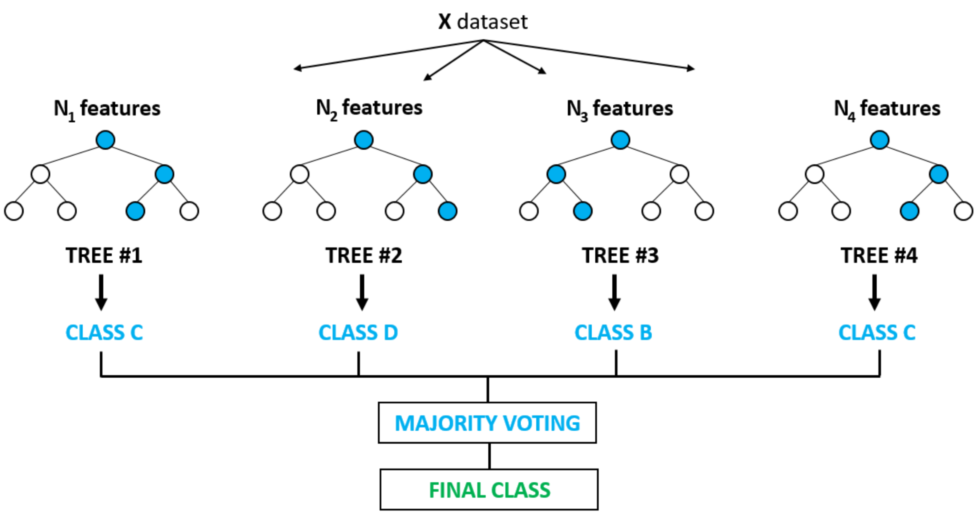

Figure 2. Random Forest classification example of four trees

2.4.2 K - Nearest Neighbors

KNN predicts the label of any new sample based on the labels of the k closest samples from the training set, returning the most common label of the neighbors as the predicted label (Shalev-Shwartz & Ben-David, 2013). “Closeness” is determined by a distance metric, like Euclidean and Minkowski. Therefore, to allow each predictive variable to contribute equally in the distance calculation, centering and scaling are suggested, to avoid any difference in the measurement scale that might affect the resulting distance calculations between samples, generating biased towards predictive variables with larger scales (Kuhn & Johnson, 2013). An example of a 5-nearest neighbor model is depicted in

Figure 3 (Kuhn & Johnson, 2013), where a classification of two new samples (denoted by the solid dot and filled triangle) is carried out. Class probability estimates for the new sample are calculated as the proportion of training set neighbors in each class. Therefore, the solid dot sample is near a combination of the two classes where is highly likely that the sample should be labeled as the first class. The other sample is surrounded mostly by neighbors of the second class, hence this one may be labeled as the second class.

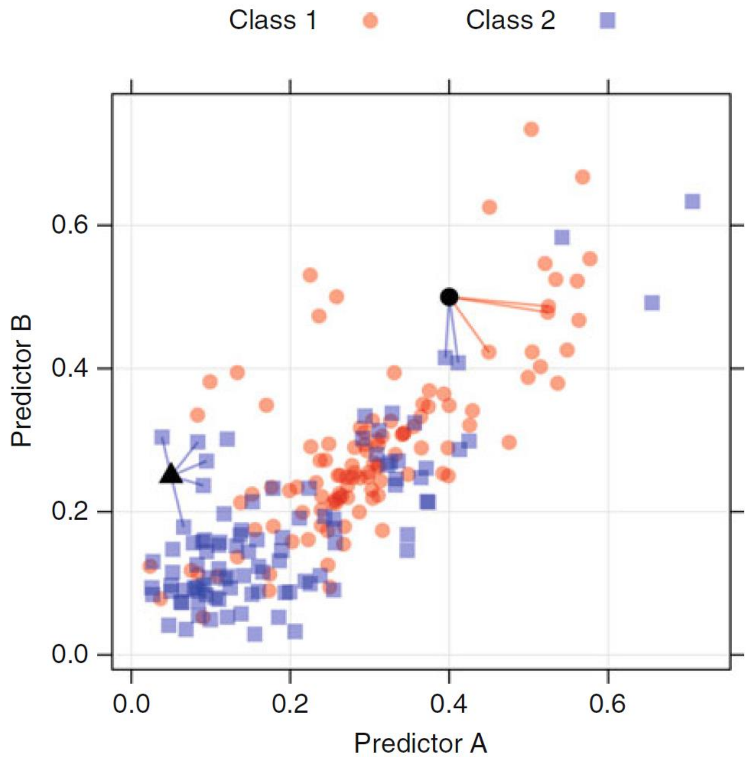

Figure 3. The K-nearest neighbor classification model.

2.4.3 Support Vector Machine

The SVM algorithm was designed for binary classification problems (Chiranjit, 2015). SVM aimed to find the hyperplane in the feature space that maximally separates the two target classes (Sarmadian et al., 2014). Therefore, multiple hyperplanes could be chosen, but the one with the maximum margin between data points of both classes would be considered as the best candidate (Chiranjit, 2015). Figure 4 (Gahukar, 2018) shows a case of linearly separable data by hyperplanes, where three possible separating hyperplanes are illustrated. It is evident that the red hyperplane (H_3) has a larger margin than the other, and is, therefore, the best candidate because of its greater generality for classified new data.

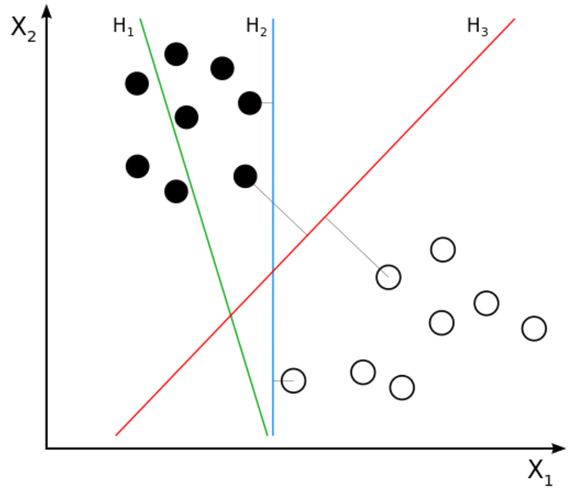

Figure 4. Linearly separable data by hyperplanes, where X_1 and X_2 are predictive variables, H_1, H_2 and H_3 are hyperplanes, the gray lines are the separation margins and the black and white points represent the two target classes

The major disadvantages of SVM are that including non-informative predicting variables can affect negatively the model and high dimensionality of the feature space increases the computational cost (Kuhn & Johnson, 2013). SVM produces very competitive results and is easy to implement because requires a minimum amount of model tuning (Chiranjit, 2015). The SVM algorithm could be a suitable alternative for the performance of land suitability analysis (Sarmadian et al., 2014).

2.4.4 Model Tuning

Classifications models have parameters that cannot be directly estimated from the data or from an analytical formula, called tuning parameters (Kuhn & Johnson, 2013). For example, in the KNN classification model, the value of K neighbors that the model used to label new samples is a tuning parameter. It is not possible to know in advance which parameter set of values will generate the best model because these parameters do not learn from the data, requiring validation strategies that allow compare different values and select the most adequate for each model (Shalev-Shwartz & Ben-David, 2013). The use of existing data to identify settings for each model parameter allowing to obtain the most realistic predictive performance is called model tuning (Kuhn & Johnson, 2013). K–Fold cross-validation method was selected to implement model tuning due to its capacity to provide an accurate measure of the true error without wasting valuable data (Shalev-Shwartz & Ben-David, 2013). In this method, the samples of the data are randomly divided into k partitions (folds) of similar sizes and each partition is used to testing a classification model trained to predict the remained samples in the other k-folds (Wong, 2015). The average performance of the hold out partitions is estimated and used to determine the final tuning parameters. A final model is trained using the selected tuning parameters on the entire data set (Kuhn & Johnson, 2013). An example of threefold cross-validation is illustrated in Figure 5 (Kuhn & Johnson, 2013).

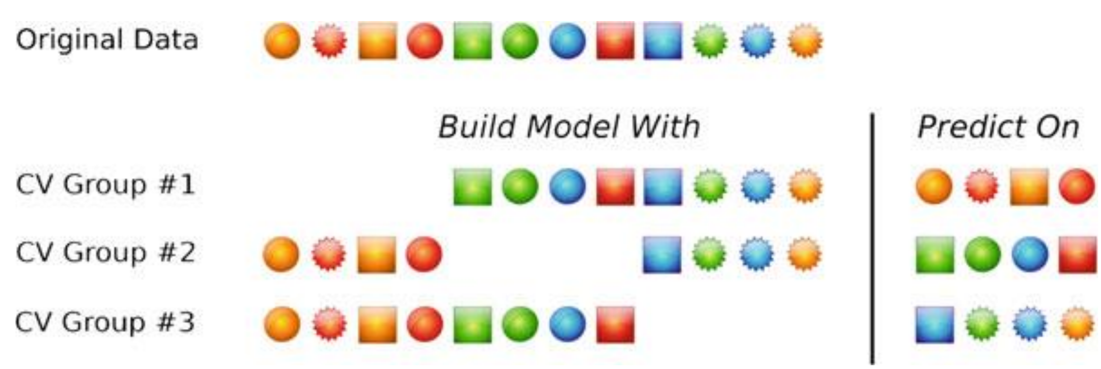

Figure 5. A schematic of threefold cross-validation. Twelve data samples are displayed as symbols and allocated in three groups. These groups are left out in turn as models are fit. Performance is estimated from each set of held-out samples and their average would be the cross-validation estimate of model performance.

2.4.5 Model Performance

In Machine Learning, model performance measurement is an important step. Different machine learning models (like regression models) usually implement Root Mean Squared Error ($RMSE$) and the coefficient of determination (R^2) metrics to evaluate the performance, but in the context of classification models, these metrics are not appropriate to assess the performance (Kuhn & Johnson, 2013). The Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) was defined for two-class problems to evaluate the class probabilities of models through a variety of thresholds, indicating how capable the model is at distinguishing between the two classes (de Figueiredo et al., 2018).

Figure 6 shows an example of the ROC curve where the AUC is a measure of the discrimination between two classes always bounded between 0 and 1 (Kuhn & Johnson, 2013). The True Positive Rate (TPR) refers to the ability to correctly identify an event as positive also called sensitivity, and Inversely, the False Positive Rate (FPR) is related to correctly identifying negative events also called specificity (de Figueiredo et al., 2018). The ROC curve plots the TPR and the FPR (one minus the specificity) against each other for each possible threshold. (Pontius & Parmentier, 2014). The model with the highest AUC value would be the best to differentiate the probabilities between the two classes.

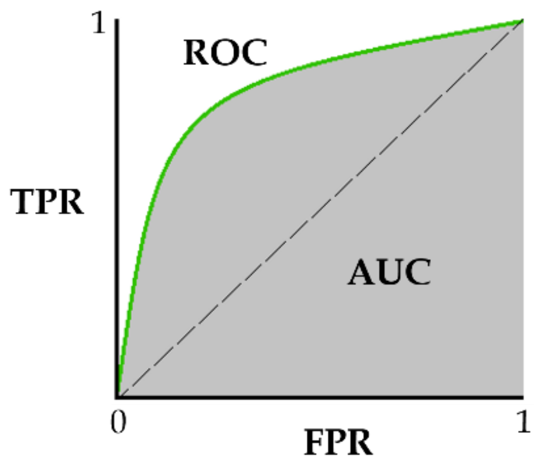

Figure 6. Graphic Example of the ROC curve

The AUC - ROC method was selected for measuring model performance in this thesis due to its simple interpretation of class probabilities in classification problems and because of its popularity in evaluating performances and assisting in the decision-making process (de Figueiredo et al., 2018). Additionally, AUC - ROC has been applied for determining the accuracy in similar studies of agriculture land Suitability Analysis using MCDM methods and machine learning techniques (Heumann et al., 2011; Montgomery & Schmidt, 2015; Parthiban & Krishnan, 2016).

3. METHODOLOGY

This chapter explains the processes developed for this research. First, the established assumptions in this thesis are mentioned in section 3.1. Then, a description of the software and hardware used is explained in section 3.2, an introduction to the study area and data used is described in section 3.3. The procedure carried out in the expert knowledge approach using the MCDM-AHP method is explained in section 3.4, including a spatial sensitivity analysis developed to evaluate the robustness of the model. The steps implemented in the data-driven approach based on machine learning techniques are explained in section 3.5. Comparison based on the relevant variables in both approaches and the performance of the models implemented is described in section 3.6. Finally, the selection of public buildings and the steps taken to the estimation of public open spaces and crop productivity is presented in section 0. An overview of the methodology implemented is depicted in Figure 7.

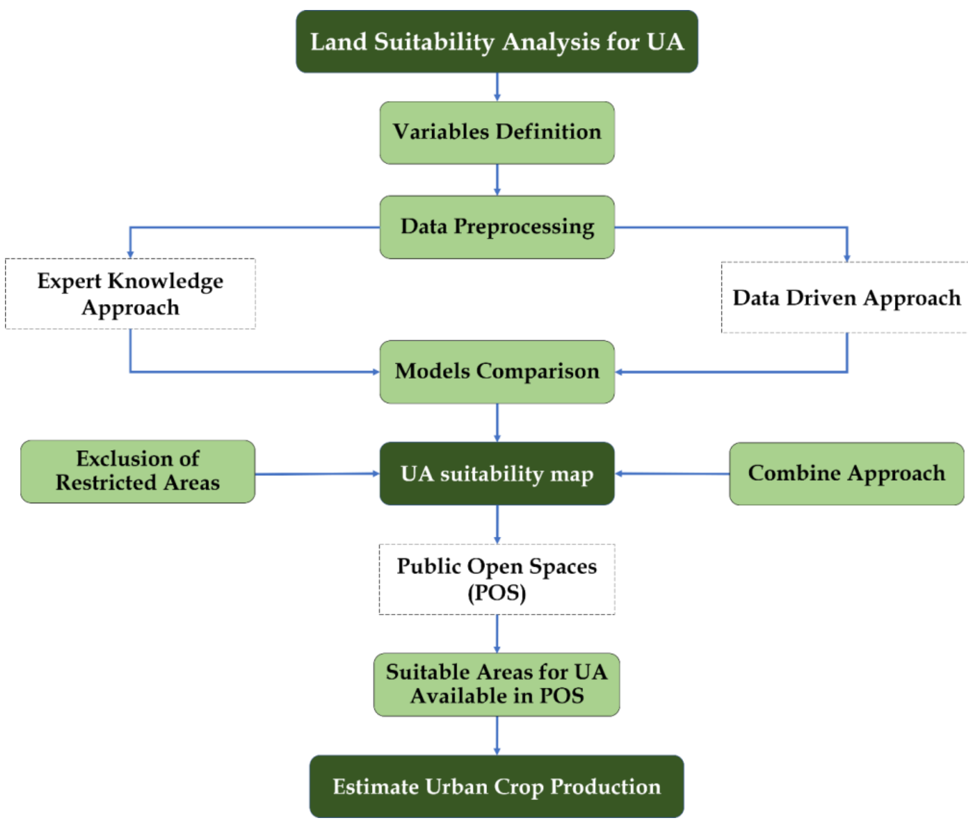

Figure 7. Flowchart of the Methodology Implemented

3.1 ASSUMPTIONS

The following assumptions were considered for the development of this thesis:

- Although non-food products can be obtained by urban agriculture activities including flowers, aromatic and medicinal herbs, ornamental plants, tree products (seed, wood, fuel, etc.) and tree seedlings (Thomas, 2014). Only vegetable production derived from urban agriculture practices, such as horticulture were considered due to the relevant contributing to food and nutrition security (Orsini et al., 2013).

- City Water can be accessed in every building or household within the city (99.86% coverage) and is suitable to use to grow crops.

- Urban agriculture activities can be developed in indoor and outdoor spaces (Thomaier et al., 2015). The scope of this thesis is oriented only urban agriculture practices in outdoor locations or open spaces.

3.2 SOFTWARE AND HARDWARE

ArcGIS Pro is a GIS application that allows visualizing spatial and attributive information, performing advanced geoprocessing analysis (ESRI, 2020). *Spatial Analyst* extension was used to perform the suitability analysis and math and conditional operations based on cell-based raster data ("Spatial analysis in ArcGIS Pro—ArcGIS Pro | ArcGIS Desktop," 2018)

R is a language for statistical computing and graphics (R Core Team, 2020). It runs on different operating systems using several packages mostly related to data analysis and visualization. The Integrated Development Environment (IDE) RStudio version 1.1.456 is used as an execution interface of the R software. R version 3.6.1 is used in combination with the following packages:

- ggplot2 (CRAN v. 3.2.1): creates elegant data visualizations using the grammar of graphics (Wickham, 2020). Mainly used for results visualization by creating plots and charts.

- caret (CRAN v. 6.0.84): provides an easy way to create predictive models based on classification and regression techniques (Kuhn, 2020).
- ROCit (CRAN v. 1.1.1): creates the Receiver operating characteristic (ROC) curve used to measure the performance of Binary Classifier with Visualization
- raster: allows reading, writing, manipulating, analyzing and modeling of raster spatial data (Hijmans, 2020)

Python is an interpreted and object-oriented programming language that due to its simplicity and readability is used in several research fields (Python.org, 2020). Python version 3.7.0 is used in combination with the following libraries and packages:

- Numpy: library for numerical computations that perform data manipulation and fast mathematical and logical operations on arrays (Van Der Walt, Colbert, & Varoquaux, 2011).
- Arcpy: used for geoprocessing analysis, spatial data manipulation and map automation with Python (Toms, 2015)
- Matplotlib: 2D plotting library for scientific publishing and interactive graphing, used to produce quality graphics (Matplotlib.org, 2020)

Software Applications	ArcGIS Pro 2.4.2
Programming Languages	Python 3.7.0, R 3.6.1
Integrated Development Environments	RStudio, PyScrpter
Data Manipulation	Numpy 1.16.3
Data Visualization	ggplot2 3.2.1, Matplotlib 3.1.0
Machine Learning Package	Caret 6.0-84
CPU	Intel(R) Core (TM) i7-6650U 2.20 GHz
Motherboard	Microsoft Surface Pro 4
RAM	16 GB DDR4

Table 3. Software and Hardware used for the research

3.3 DATA AND STUDY AREA

3.3.1 Study Area

This study focuses on Bogotá city, the capital and the biggest city of Colombia with a latitude of 4° 36' 34.96" north, and longitude of -74° 04' 54.30" west and an altitude of 2.640 m above sea level. The study area is integrated by the urban and urban sprawl areas defined by the urban planning department of the city. Figure 8 shows the location of the study area with an extension of 40.716 Hectares.

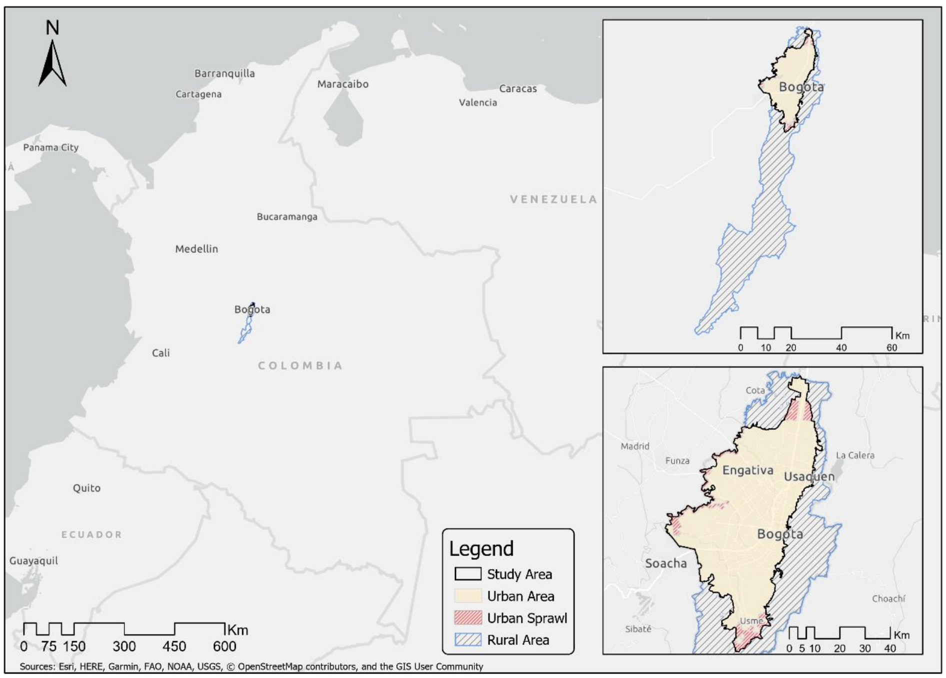

Figure 8. Study Area

3.3.2 Data Description

The data used in this thesis is a compendium of different sources obtained by local and governmental entities of the city, which may represent the physical, environmental and socioeconomic components of the study area. Additionally, significant variables for the development of crops and plants in an urban environment were considered based on the received feedback of local experts and similar land suitability studies (Aldababseh et al., 2018; Feizizadeh & Blaschke, 2013; McClintock et al., 2013; Setiawan et al., 2014; Spataru, Faggian, & Sposito, 2018; Thornton, Momoh, & Tengbe, 2012; Uy & Nakagoshi, 2008a).

The term 'expert' used in this study, refers to the group of people that due to their academic or professional experience related to urban agriculture contributed with their knowledge in the assessment of the procedures implemented in this thesis. This group is mainly composed of 8 members as shown below:

- The coordinator of the urban agriculture department of the Bogota Botanical Garden
- Two professors with academic knowledge in sustainable agriculture and organic farming
- Two soil scientists with professional experience in agronomy
- One environmental scientist with professional experience of GIS applied to agricultural studies
- Two urban farmers with local knowledge about urban crop production

Although the temperature and different soil properties are vital elements for the development, growth, and productivity of crops in agriculture, being considered relevant variables included in several land suitability analysis; these were not included in this research based on the following criteria:

- Unlike traditional agriculture, urban agriculture is not completely dependent on soil for its development, since it can be created artificially using different methods based on the mixing of organic and inorganic residues to create natural fertilizers that provide the necessary nutrients for the crops.
- The temperature was not considered because this research was not focused on the analysis of a particular agricultural crop, but the identification of potential sites for the specific development of urban cold climate crops native to the study area (Bogota Botanical Garden, 2007).

In total fourteen variables were selected and grouped in two main components based on interpretation of literature reviews of internal and external references, availability of data and expert knowledge. Table 4 and Table 5 provide information related to the variables selected for this research, as well as a short definition of each variable, their units, main source-year of the data and their relevance of implementation or study for the UA.

The Bogota Botanical Garden supplied the location in Shapefile format of the current urban orchards in the city (Figure 9). In total, for the study area, there is a record of 202 urban orchards, of which 106 are in private spaces (mainly residential units) and 96 in public spaces (universities, schools, kindergartens, medical centers, etc.). About the type of organization, 79 orchards are managed by communities, 46 by institutions, 23 by schools and 52 by families (Figure 10).

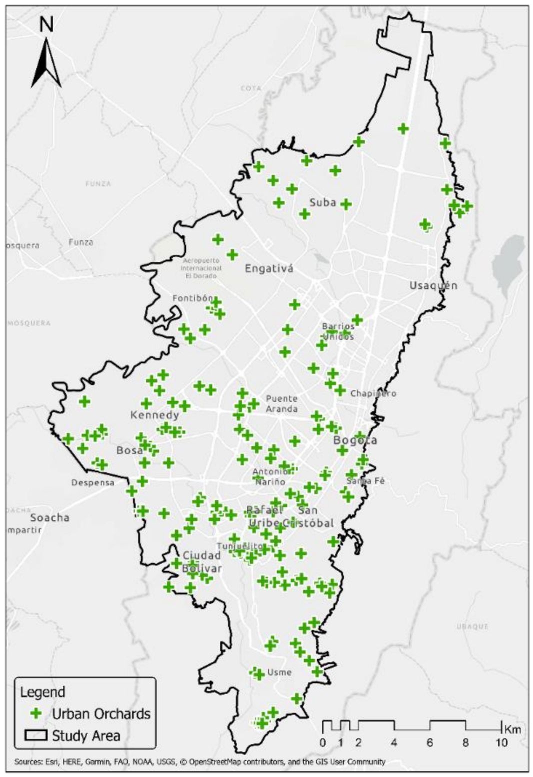

Figure 9. Locations of urban orchards in the study area

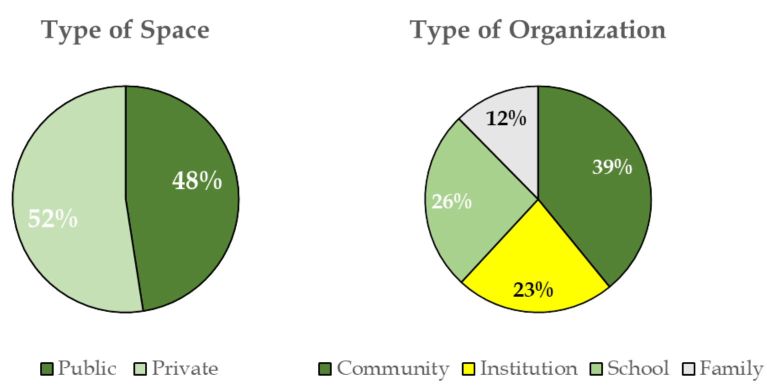

Figure 10. Type of space and organization of the urban orchards

Table 4. Variables description for the physical and environmental component

Component	Variable (definition)	Dimensional Unit	Relevance for AU	Source
Physical and Environmental	**Sunlight**: also called sunshine, is the number of hours in which the light and energy coming from the Sun reaches the earth's surface directly;	hours/day (3.1 – 4.42)	Solar energy is used to apply drying techniques on plants, stems, roots or fruits, to preserve the tissues and their nutritional, medicinal or aromatic properties. Solar radiation, or sunshine, is the primary source of light and heat energy for plants and crops.	IDEAM 2018
	Rainfall: the amount of water that falls on the earth's surface in liquid or solid form. In practical terms, it is the average amount of rain that falls in a given area and contributes to the water requirement of crops	Millimeters (mm) (215.5 -1143.8)	Rainwater harvesting (roofs and terraces) for storage in tanks or containers and use in crops.	SDA 2018
	Slope: identifies the difference in gradient between two relief forms. Relationship between the horizontal distance and the altitude between two points.	Percentage (0 – 100%)	The ground can be flat or moderately sloped, for which it is recommended to sow in furrows in the opposite direction to the slope.	UAECD 2018
	Aspect: can be described as the direction of the slope. Identifies the direction of the downward slope of the maximum rate of change in value from each cell to its neighbors.	Degrees (0 -360)	The location of crops oriented north to south is preferred to guarantee constant light.	UAECD 2018
	Hillshade: represents the presence of shade due to the relief and height of buildings	Qualitative	On open spaces such as terraces and patios, sunlight is generally not affected, because there is no interference or blockage in the sun's exposure; in areas surrounded by buildings, sunlight can be affected due to the interference caused by the shadow of the buildings.	UAECD 2018
	Roads Distance: Set of lines that represents the road network	Distance in meters (m)	facilitating maintenance work and purchase of materials necessary for the implementation of crops (beds and containers). the location must have good access to facilitate transport and movement.	UAECD 2018

Component	Variable (definition)	Dimensional Unit	Relevance for AU	Source
Social and Economic	**Public Space Indicator (PSI):** establishes the relationship between Effective Public Space (public space of a permanent nature, made up of green areas, parks, squares, and small squares) and the population	People/m2 (< 3; 3 – 6; 6 – 9; 9 – 15; > 15)	Representative parks, green areas, or small squares in public spaces that could represent potential places for the establishment of AU practices.	DADEP 2018
	Population Density: the ratio of the number of people per hectare	People/ha (0 – 633.4)	Identify areas where a greater number of people located in residential properties and housing units can benefit from implementing AU practices	SDP 2018
	Residential Density: the ratio of the number of dwelling units per hectare	Dwelling units /ha (0 -114.7)		UAECD 2018
	Dependency Index: relationship between the dependent people, (<15 and >64 years), and the population in working age (≥15 and 64 years). The data shows the ratio of dependents per 100 persons of working age.	Index (31.58 – 50.12)		DANE 2017
	Multidimensional Poverty Index (MPI): identifies multiple deprivations at the household and individual level in health, education, and standard of living	Index (0.6 – 10.9)	Identify zones with low socioeconomic levels, extreme poverty, nutritional and food problems within the study area, to focus efforts on the establishment of possible scenarios that contribute to improving food and nutritional security by implementing AU practices	DANE 2017
	Monetary Poverty (MPv): percentage of the population with income below to the minimum monthly income defined as necessary to meet their basic needs	Percentage (3.06 – 33.85)		DANE 2017
	Unemployment Rate (UmpRate): relationship between the unemployment people and the working population	Rate (4.3 – 13.55)		DANE 2017
	Residential Areas with Low Socioeconomic Level (RALSE): weighting of properties for residential use classified with low economic levels according to their socioeconomic stratification	Index (0 – 0.6)		DANE 2017

Table 5. Variables description for the social and economic component

- UAECD – Unidad Administrativa Especial de Catastro Distrital; DANE - Departamento Administrativo Nacional de Estadística; SDP - Secretaría Distrital de Planeación; DADEP - Departamento Administrativo de la Defensoría del Espacio Público; SDA - Secretaría Distrital de Ambiente; IDEAM - Instituto de Hidrología, Meteorología y Estudios Ambientales

3.3.3 Data Preparation

Aspect and Slope layers were obtained from a DTM of 5m resolution. The hillshade layer required a more extensive preprocessing. It was necessary to identify areas would be most affected by shadows during the day due to the presence of buildings around. To do this the following steps were implemented:

- Buildings height was estimated by selecting the dwelling units and multiplying the number of floors by 2.5 m (average height of a residential floor in the city of Bogota) and for the rest of the buildings, a value of 3m was used for height estimation.
- A raster layer was created using the estimated building's height and added to the DTM cell values.
- Hillshade maps were created for three periods of time in the day, the morning time (8 am), midday (12 pm) and afternoon (4 pm) using the average values of the azimuth and elevation of the first day of each month for 2018.
- A final hillshade composes map was created for the study area identifying areas that would not be affected by shadows during the day and areas that would be affected in one, two or three periods of time during the day.

Finally, all variables were adjusted to the extent of the study area and converted into raster layers at a spatial resolution of 5m which is the coarsest resolution of the available spatial layers. Figure shows the result of the variables after the processes mentioned in this section.

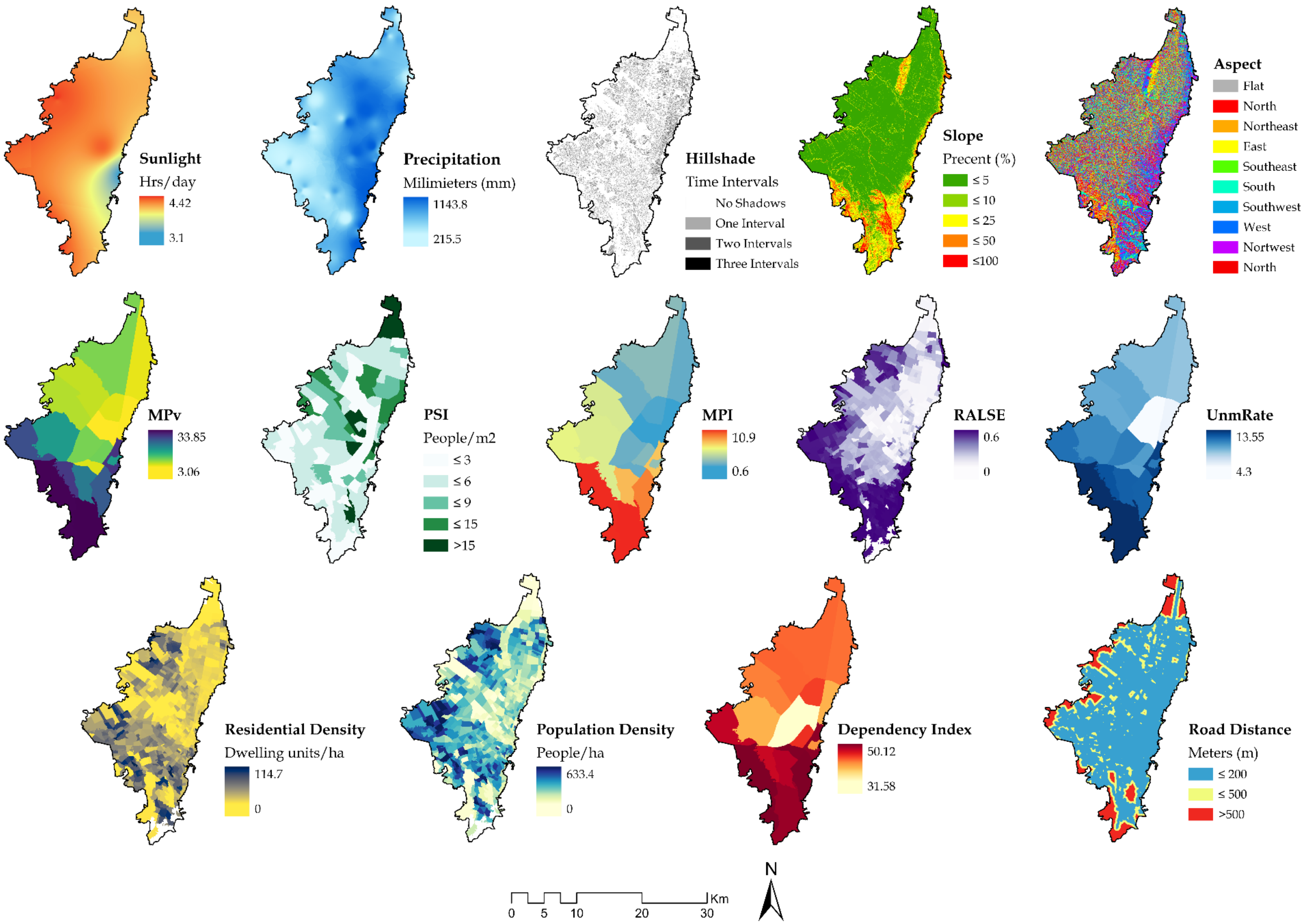

Figure 11. Variables selected for the study area. MPI (Multidimensional Poverty Index); UnmRate (Unemployment Rate); PSI (Public Space Indicator); Mpv (Monetary Poverty); RALSE (Residential Areas with Low Socioeconomic Level)

3.4 EXPERT KNOWLEDGE APPROACH: MCDM - AHP METHOD

This approach aims to define land suitability for urban agriculture in the study area, combining the potential of spatial data manipulation and analysis provide by GIS techniques with MCDM using the AHP method. Knowledge, support, and feedback from the experts are fundamental during the development of this method, in which, most of the procedures require their participation except for the steps developed within the GIS environment as can be seen in Figure 12.

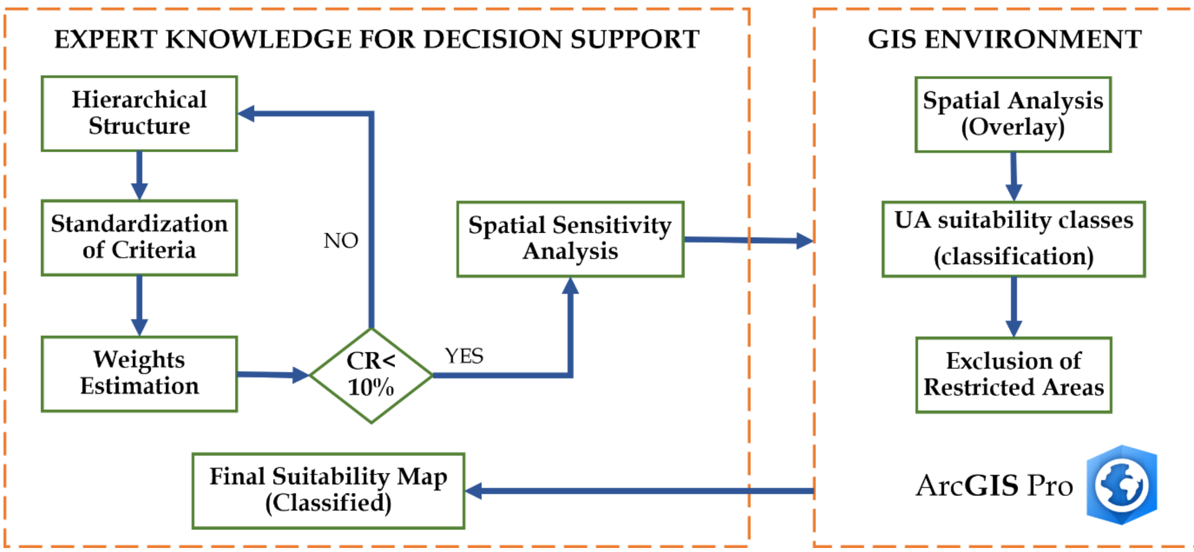

Figure 12. Flowchart of the land suitability map for urban agriculture based on expert knowledge approach

3.4.1 Defining Land Suitability Classification for Urban Agriculture

Land evaluation is a process that allows the identification and assessment of specific uses that are adapted to specific conditions of the land assessed (FAO, 2007). Although the FAO (Food and Agriculture Organization) system presents some limitations because of its orientation mainly on the physical aspect, it has been the most widely used procedure to address local, regional land management. FAO framework proposes a set of qualities and characteristics to be used in the land evaluation process (called in this research as criteria and variables, respectively) where the number is flexible and usually is determined by the objectives of the study, the scope of the research and the data available

(FAO, 2007). The classification system used in this thesis was inspired by the FAO approach dedicated to sustainable agriculture, where suitability is a measure of how well the qualities of a land unit match the requirements of a particular form of land use. For this research, the land suitability for urban agriculture was classified into three main categories ranging from most or highly suitable to marginally suitable based on the contribution to the alleviation of urban poverty by increasing food security and nutrition in the study area Table 6.

Suitability Class	Value	Description
Highly suitable (A1)	3	Land having no significant or with minor limitations (in the socioeconomic, physic or environmental components) to implement UA practices
Moderately Suitable (A2)	2	Land with moderate limitations (in any component) for implementations of UA
Marginally suitable (A3)	1	Land with marginal limitations (in more than one component) for implementations of UA

Table 6. Land Suitability Classification and Definition Used for Urban Agriculture

3.4.2 Building a Hierarchical Structure

A decomposition process was built in a hierarchical structure where the overall objective was to obtain a land suitability map for UA, including the criteria and variables used to define land suitability. In total, five main criteria were defined: climate conditions, topography, urban density, urban accessibility and socioeconomic. Subsequently, the decomposition continues to define the variables under each one of these five main criteria. The resulting hierarchal structure is shown in Figure 13.

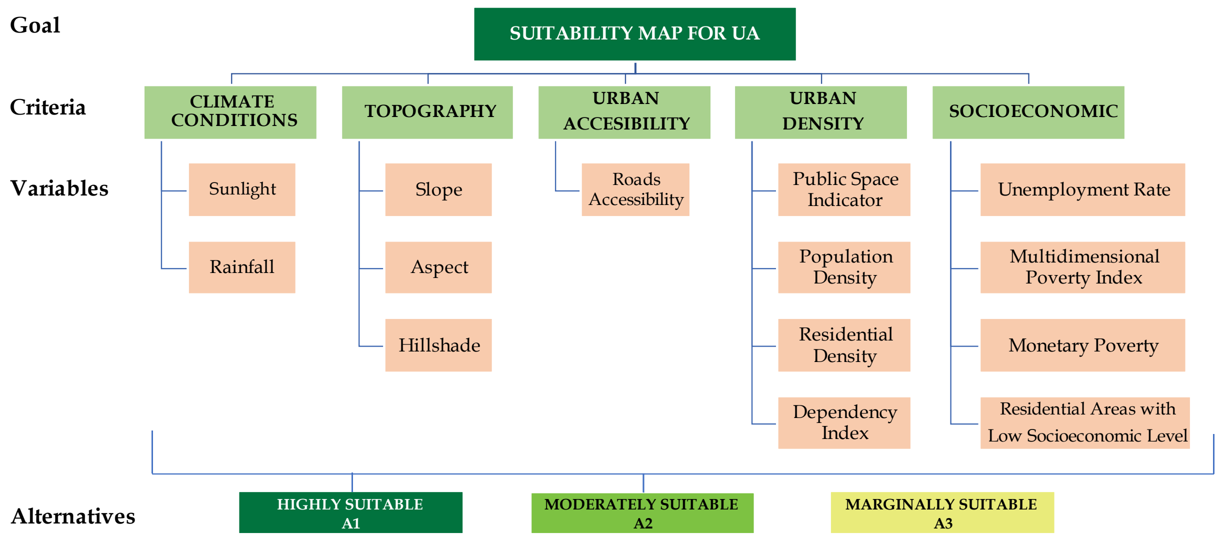

Figure 13. Hierarchal structure for defined land suitability for UA

The purpose of this procedure is to identify certain factors that affect the suitability of the land, which can be quantified in a specific range of values. For example, it is not easy to estimate numerically how climate conditions affect land suitability in general. However, when it is decomposed in a decision tree into sunlight and rainfall variables, each of these variables can be easily quantified to provide a more feasible approximation (Aldababseh et al., 2018). An example of a decision tree for the climate conditions criteria is depicted in Table 7. Decision tree tables for the other criteria can be found in **Appendix A**.

Sunlight (hours/day)	Precipitation (mm)	Suitability Class
⩾ 4	⩾ 500	A1
	< 500	A2
⩾ 3.5 - < 4	⩾ 500	A2
	< 500	A3
< 3.5	⩾ 500	A3
	< 500	A3

Table 7. Climate conditions criteria decision tree

3.4.3 Standardization of the Criteria

Evaluation criteria in land suitability analysis are represented by qualitative values or classes, indicating the degree of suitability which will be represented in the final suitability map (Prakash TN, 2003). The variables selected for the study were classified in suitability classes (A1, A2, and A3) according to the values defined in Table 8 using the reclassify tool located in ArcGIS Pro. Based on the decision trees defined in the hierarchical structure, the variables were combined using the combine tool located in ArcGIS Pro to conformed the five main criteria. Then, suitability classes were rated to define their relative importance in the main criteria and to establish numerically how these would contribute to the final suitability map. Therefore, values in Table 6 were used to rate the classes for the five main criteria. Figure 14 shows an example of the standardization process for the climate conditions criteria.

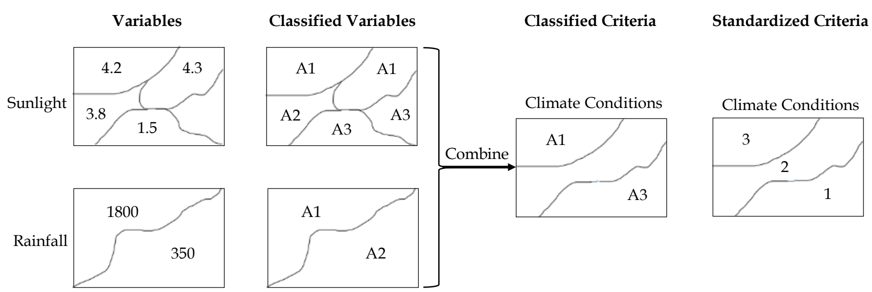

Figure 14. Standardization process for climate conditions criteria

3.4.4 Assessing the Weights and Consistency

A pairwise comparison matrix where the five main criteria were compared with themselves was constructed using Saaty's scale measurement (Table 1) to assign the relative importance of one criterion over one another. Criterion weights were estimated using the eigenvector corresponding to the largest eigenvalue of the matrix (λ_{max}) and then normalizing the sum of the components. With the weights defined the next step is to check the consistency of the pairwise comparison matrix obtained. Using λ_{max} estimated previously and a Random Index equivalent to the five criteria, the Consistency Index (CI) and subsequently the Consistency Ratio (CR) were estimated to validate the consistency of the matrix, indicating if the values for criteria comparison were assigned randomly.

	Criteria	Variable	Dimensional Unit	UA Suitability Class: Highly Suitable (A1)	Moderately Suitable (A2)	Marginally Suitable (A3)
Physical & Environmental	Climate Conditions	Sunlight	hours/day (3.1 – 4.42)	≥ 4	≥ 3.5 - < 4	< 3.5
		Rainfall	Millimeters (mm) (215.5 -1143.8)	≥ 500	<500	-
	Topography	Slope	Percentage (0 – 100%)	≤ 25	> 25 - ≤ 50	> 50
		Aspect	Degrees (0 -360)	North (0–22.5; 337.5–360) South (112.5–247.5)	West 22.5–112.5 - East 247.5–337.5	-
		Hillshade	Qualitative (Presence of shadows)	No Shadows	Once or Twice times per day	Three times per day
	Urban Accessibility	Roads Distance	Meters (m)	≤ 200	> 200 - ≤ 500	> 500
Social & Economic	Urban Density	Public Space Indicator (PSI)	People/m²	≥ 6	≥ 3 - < 6	< 3
		Population Density	People/ha	≥ 200	≥ 100 - < 200	< 100
		Residential Density	Dwelling units/ha	≥ 25	≥ 5 - < 25	< 5
		Dependency Index	Index (31.58 – 50.12)	≥ 45	≥ 40 - < 45	< 40
	Socioeconomic	Multidimensional Poverty Index (MPI)	Index (0.6 – 10.9)	≥ 5	≥ 3 - < 5	< 3
		Monetary Poverty (MPv)	Percentage (3.06 – 33.85)	≥ 20	≥ 10 - < 20	< 10
		Unemployment Rate (UmpRate)	Rate (4.3 – 13.55)	≥ 10	≥ 5 - < 10	< 5
		Residential Areas with Low Socioeconomic Level (RALSE)	Index (0 – 0.6)	≥ 0.4	≥ 0.2 - < 0.4	< 0.2

Table 8. Hierarchical structure and classification of variables using suitability classes defined for UA

3.4.5 Spatial Sensitivity Analysis

Once the criterion weights were estimated and the consistency was validated, the variables and the standardized criteria were integrated as information layers in format raster into a GIS environment using ArcGIS Pro. A spatial sensitivity analysis was carried out using the OAT method to assess the uncertainties in the criteria weights obtained from the pairwise comparison matrix in the AHP method and determine the robustness of the results. Figure 15 summarizes the steps implemented in the spatial sensitivity analysis using the One at a Time (OAT) method.

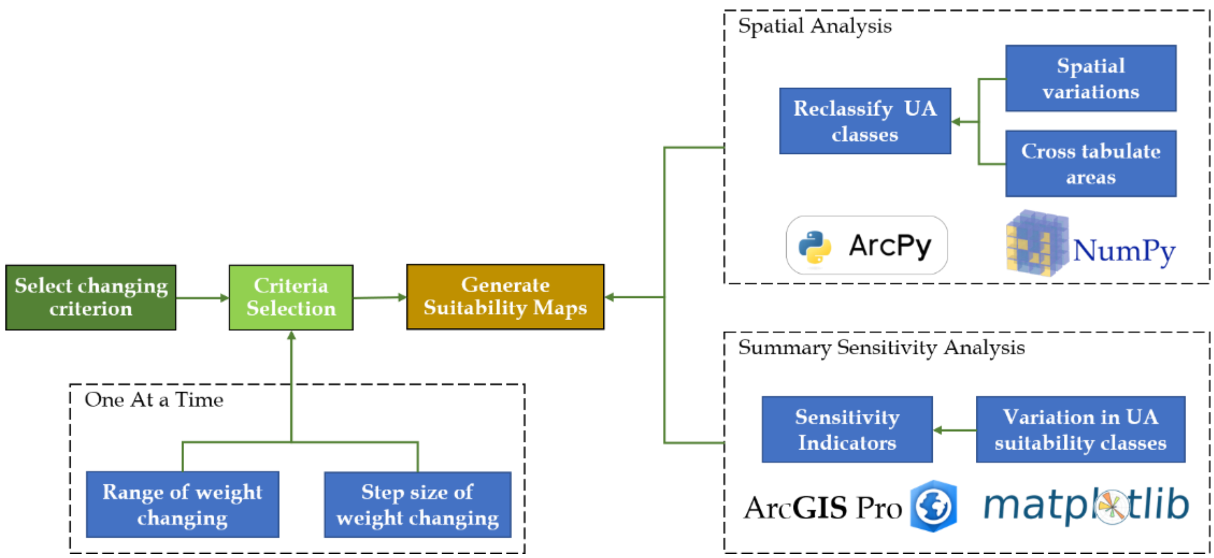

Figure 15. Flowchart of the spatial sensitivity analysis

The main input for the OAT method was the criteria weights obtained as a result of the MCDM - AHP method. Additionally, the OAT method requires setting two parameters, the range of the weight changing (RWC) and the step size of the particular weight changes (SWC). For this research, an RWC of ±20% and an SWC of ±1% were applied over all the five main criteria. Each criteria weight was modified ±1% while the values of the other four criteria were adjusted assuring that the sum of all criteria weights must be equal to 1 (Equation 5). This process was repeated until the variation of the criteria weight selected would be equal to ±20%. Tables per criteria with these weight variations were created and stored in the GIS environment. A Python algorithm was created to use the existing functions and methods available in the ArcPy library for the spatial analyst and Numpy library for data manipulation. This arcpy script read the criteria tables with the weight variations created previously, and using a Weighted Using Overlay function the standardized criteria were multiply by their correspondent SWC (from ±1 to ±20), generating unified land suitability maps that subsequently were reclassified in the UA suitability classes defined in Table 6.An example of this procedure is depicted in Figure 16, where w is the SWC of the corresponding criteria.

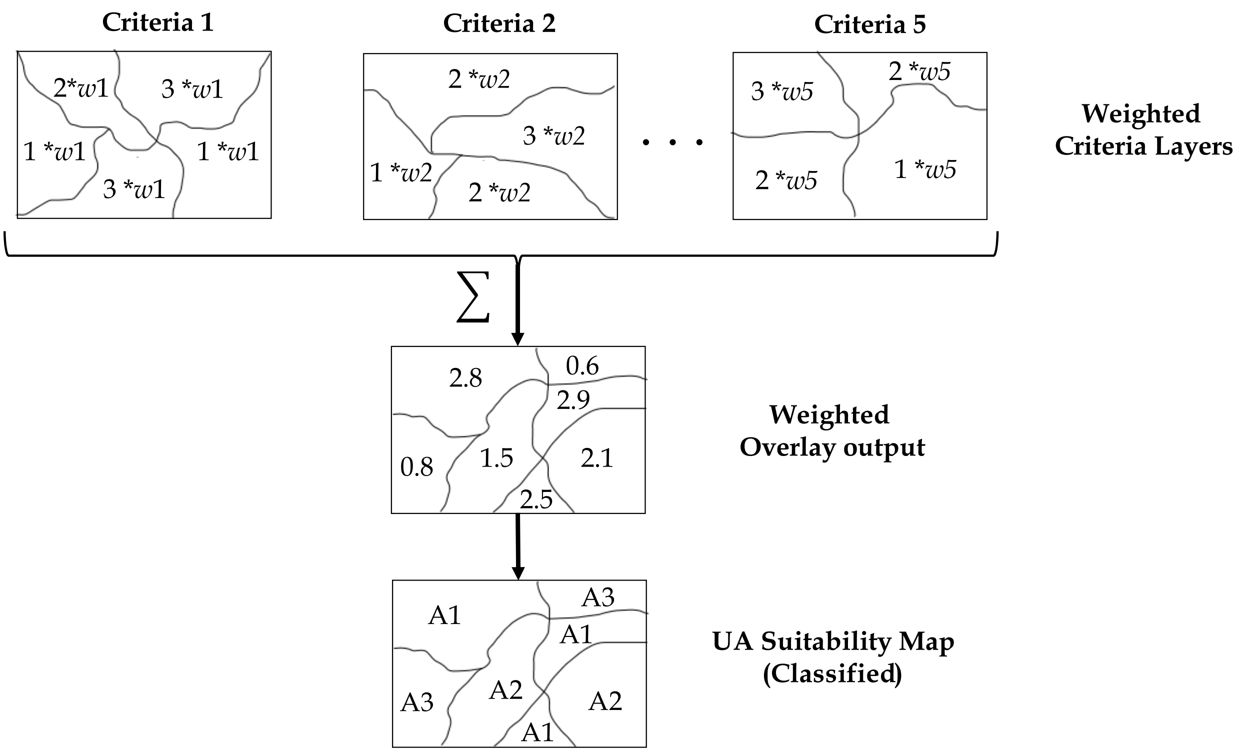

Figure 16. Weighted overlay process used to obtain the suitability map for UA

Sensitivity indicators were created using cross-tabulate areas between the baseline suitability map (0% weight change) and each new suitability map generated, identifying the changes in the cell values that shifted between suitability classes and estimating the percentage of variation respect the baseline suitability values. Tables and plot graphs (using matplotlib library) summarizing weight variations and sensitivity indicators for each criterion were created, and a map for visualization the maximum changes in evaluation results for the five main criteria was generated using ArcGIS Pro. Results allowed experts to assess the robustness of the MCDM – AHP model implemented and defined the final criteria weights that would be used to generate the UA suitability map in the study area.

3.5 DATA-DRIVEN APPROACH: MACHINE LEARNING TECHNIQUES

There are different packages and modeling functions developed for model training and prediction based on machine learning in R, which could be overwhelming due to the selection of which algorithm belongs to which package and the different designed, syntax, inputs and outputs (Kuhn, 2019). The Caret (Classification and Regression Training) package was designed to build, test and compare different machine learning models in a more efficient and automated workflow, being the consistent modeling syntax one of the most powerful aspects of this package facilitating the work with different functions and models. Unlike the expert knowledge approach, where a set of criteria weights is estimated based on subjective human judgments to generate a UA suitability map. This approach aims to learn from current locations where UA practices are being implemented (represented as urban orchards) in the study area, by machine learning techniques to obtain weights automatically and objectively that later are used to generate a UA suitability map. Machine learning models applied in this approach were created within the R studio environment using the caret package. The main steps implemented in this approach could be seen in Figure 17, where ***Data preparation and preprocessing*** is oriented to set and arrange the variables and formats that require modeling functions, ***Model training and tuning*** where the machine learning models are built base on the required parameters, ***Model prediction*** UA suitability maps are created base on the predicted class for the study area and in ***Model performance,*** the measure used to validate the model performance is described. Summarize the main steps developed in this approach.

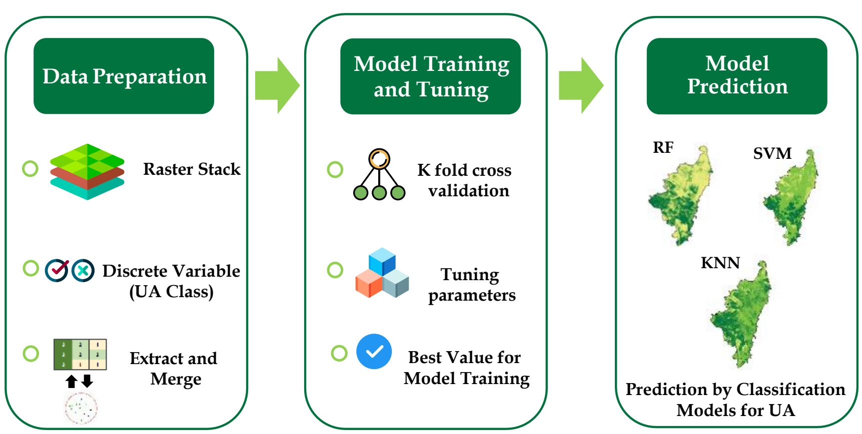

Figure 17. flowchart of the data-driven approach

3.5.1 Data Preparation and Preprocessing

To evaluate and being able to compare with the results obtained in the expert knowledge approach (section 3.4) is necessary to work with the same variables handled within the criteria defined (Figure 13). Therefore, a raster stack containing these variables was created and used as a predictor (variables used to predict the target or response variable) and the location of the urban orchards was used to define the response variable. Urban orchards represent the presence of UA activities in the study area. However, for classification models, class probabilities provide further insights into model predictions than the simple class value (Kuhn & Johnson, 2013). Based on this, a discrete variable called *UA class* of two classes was created to represent the presence and absence of UA activities in the study area. For the first class, locations of the urban orchards were selected and labeled as '*Yes*' to indicate the presence of UA activities within the study area. Therefore, the same number of urban orchards was used to create random points (pseudo-absence data) within the study area. For classification and machine learning techniques, a random selection of geographically pseudo-absence data had the most significant impact on model accuracy (Barbet-Massin, Jiguet, Albert, & Thuiller, 2012). The random points created were assigned to the second class of the discrete variable *UA class* and labeled as '*No*' to represent the absence of UA activities within the study area. The *UA class* variable created was used to represent the response variable in the classification models. Figure 18 shows the categories defined in the *UA class* variable.

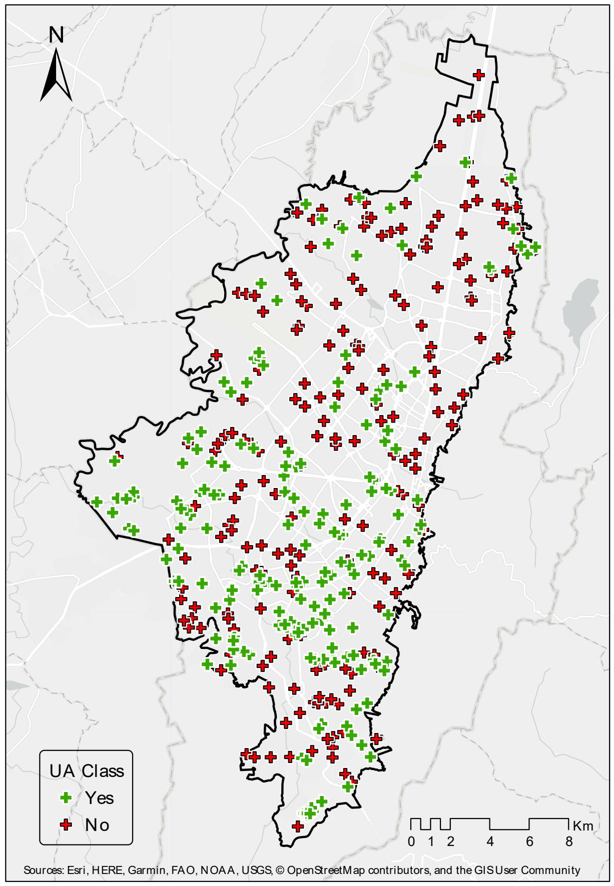

Figure 18. Categories of the response variable (*UA class*) defined to represent the presence (Yes) and absence (No) of UA activities within the study area

To train the classification models selected between the predictor variables and the response variable, an extraction process of the pixel values of the raster stack (previously created) that overlay with the locations of the *UA Class* variable was carried out. The resulting pixel values were merged with the information of the *UA class* variable and this output was stored in a data frame. This data frame contains information about the predictor variables for each location of the response variable in the study area.

3.5.2 Model training and tuning

K fold cross-validation was implemented for model tuning and evaluation of the performance of the classification models selected. Table 9 shows the classification models selected with their corresponding parameters. The selection of K is arbitrary, but usually, a selection of K=5 or K = 10 is implemented (Kuhn & Johnson, 2013).

Model	Parameters	Label
Random Forest	mtry	Randomly selected predictors
Support Vector Machine	sigma	Sigma
	C	Cost
K nearest neighbors	K	Number of neighbors

Table 9. Parameters of the machine learning models selected for the study

The best tuning parameter values for the classification models selected were obtained by a 10-fold cross-validation process. Subsequently, the three models (RF, SVM, and KNN) were trained using the selected tuning parameters on the entire data set contained in the *UA class* variable.

3.5.3 Model Prediction

Using the data frame with information of the predictors and the response variable, and the models selected trained, binary predictions of the classes (*Yes/No*) were created for each model. The values obtained by the prediction of the class *Yes* were used to represent the UA suitability in the study area. These predictive values were in the form of a probability (between 0 and 1), hence, to be able to compare the results with the expert knowledge approach, it was necessary to reclassify these values in the UA suitability classes defined in Table 6. The probability values of the class *Yes* were classified based on the ranges defined in Table 10 and UA suitability maps were generated for each model.

Probability Values	Suitability Class
0 – 0.3	A3
0.3 – 0.6	A2
0.6 - 1	A1

Table 10. Suitability classes defined for the predicted UA class values

3.6 MODELS COMPARISON

The classification models (RF, SVM, and KNN) implemented in the data-driven approach were compared with the MCDM-AHP model applied in the expert knowledge approach statistically based on the model performance and visually based on the relevant variables in each approach.

3.6.1 Models Performance

The Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) curve was applied to evaluate the class probabilities for the *UA Class* variable in the models implemented, identifying how capable was each model at distinguishing between the two classes (*Yes/No*). ROC curves were constructed for each model and the AUC value was used as a performance measure to compare both approaches.

3.6.2 Visualization of the Relevant Variables

Criteria weights obtained by the MCDM-AHP method were used to represent the variable importance for the expert knowledge approach and the classification model with the highest accuracy based on AUC results was selected to represent the variable importance for the data-driven approach. Bar charts of the variable importance for each approach were created and plotted side by side for comparison.

3.7 LAND SUITABILITY MAPS FOR URBAN AGRICULTURE

There are areas that due to their location or participation in the ecosystems and biodiversity of the city are considered as protected or relevant for ecological conservation. Therefore, it was necessary to exclude them from the analysis because according to local authorities any kind of activity (including UA) cannot be carried out in these areas. Additionally, bridges, roads, bicycle paths, and pedestrian areas were also excluded from the analysis. With the exclusion of these areas, three final land suitability maps for UA practices in the study area were created using ArcGIS Pro: two maps representing the result from the methods applied in each approach and a combined map representing the land suitable areas for UA in common between the two approaches implemented. These common areas were obtained using a cross-tabulate process in ArcGIS Pro. In Addition, graphics and tables summarizing the UA suitability areas (ha) were created.

3.8 PUBLIC OPEN SPACES AND URBAN CROPS

For this study, *public spaces* were considered as public and private parcels of land managed by local or government entities (e.g., city hall or courthouses) or parcels where community services are provided (e.g., universities, churches or parks) in which it is more feasible to implement UA practices either by state policies or by self-interest. ***Public Open Spaces*** are areas of public spaces without built or constructed areas. Figure 19 summarizes the steps implemented for the selection of the public open spaces in the study area. ArcGIS Pro software was used mainly for spatial data manipulation and visualization. Public open spaces were classified into seven public categories based on the type of activity or building that manages its area or space (Figure 20).

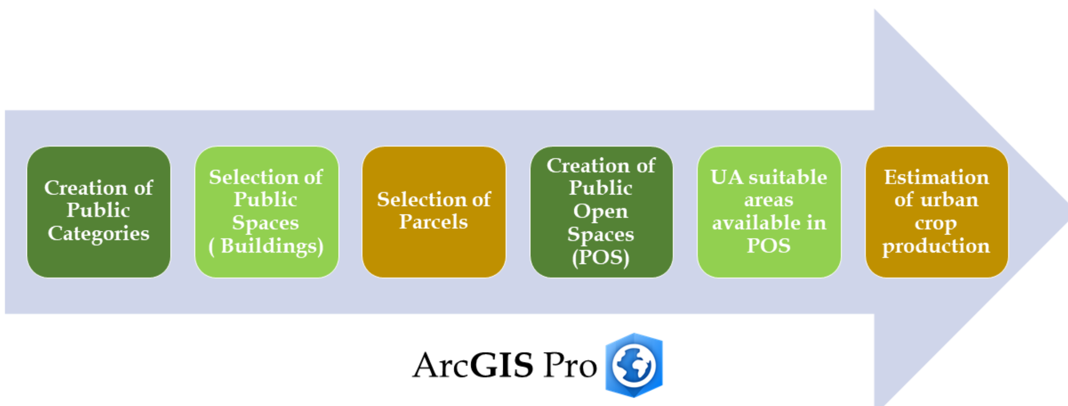

Figure 19. Public Open Spaces (POS) and Urban Crops Estimation Methodology

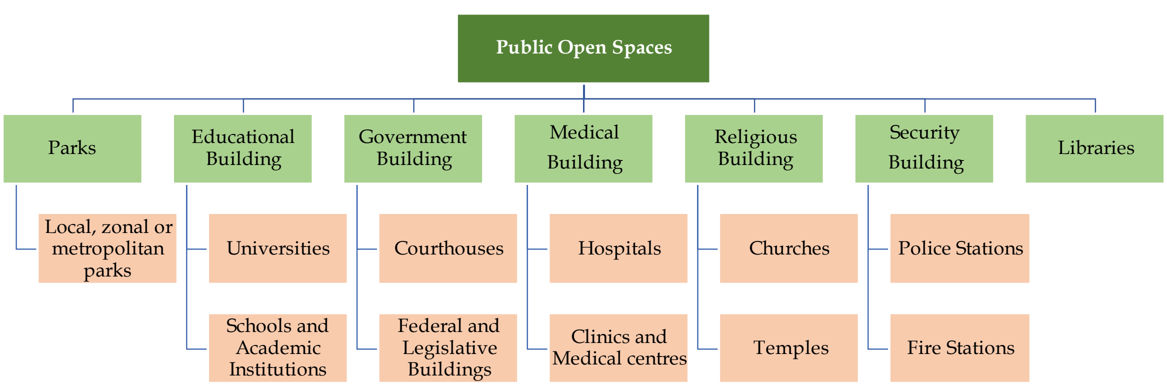

Figure 20. Public categories defined for Open Public Spaces

A selection of parcels that overlay with the public categories locations was carried out. The public open spaces layer was created removing the areas of the public categories from the selected parcels. An example of this process is shown in Figure 21, where an educational building (red polygon) was selected and then the parcel that overlaid with this building (yellow polygon). Subsequently, the public open space (blue area) is the resulting area of the extraction of the building area from the parcel.

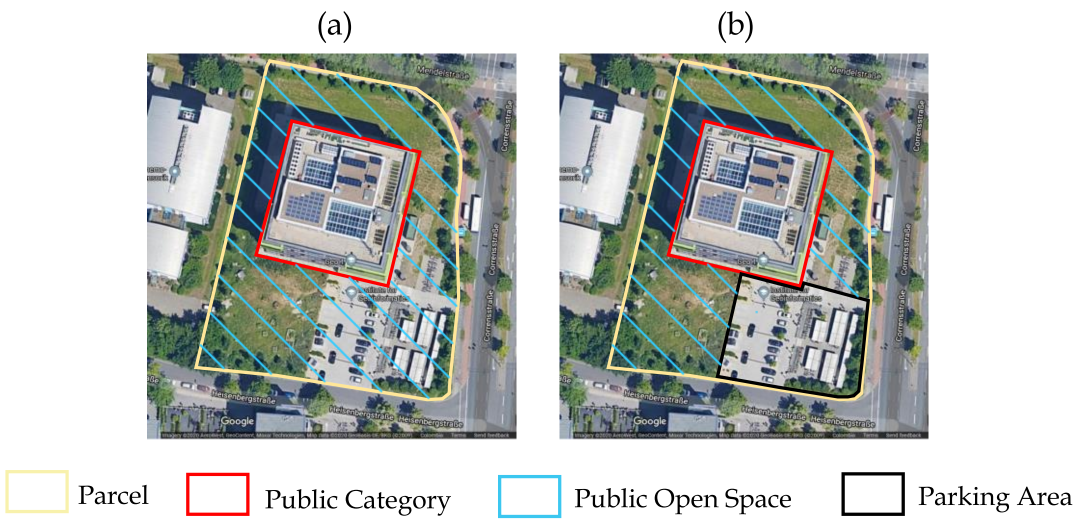

Figure 21. Public Open Spaces example. (a) selection of public categories and parcels used to create the public open spaces layer. (b)

Land suitability maps for UA obtained from the expert knowledge and data-driven approaches were overlapping with the public open spaces layer. From this result, a table and plots summarizing suitability classes area grouped by public categories were created. Likely, not all the available areas estimated in open public spaces will be implemented for the development of UA practices because these areas have or shared some other land uses. For example, not all 100% of the area estimated as public open space in Figure 21 could be used for UA activities, because there is an area inside that space destined for parking (black polygon). For this reason, three possible scenarios for the development of UA practices in public open spaces were suggested. The first scenario states that 10% of the areas available in public open spaces would be used for future UA activities, the second and third scenarios vary only in the percentage of area implemented with 30% and 50% respectively. Additionally, experts suggested that approximately 50%of the areas destined for UA activities are used for crop production. Some areas should be considered for the mobility of urban farmers, spaces between crops, places for storage and maintenance, etc. Therefore, this percentage was also considered within the scenarios established. It should be pointed out that the harvested yield of several crops grown using UA techniques is much higher compared to the yield of traditional agriculture due to the intensive use of the land in small spaces (Cantor, 2010). The average harvested yield for five of the most popular crops grown using UA in the study area based on local urban farmers and expert knowledge is shown in

Table 11.(Cantor, 2010). Using the information of the land suitability areas for UA available in public open spaces obtained from the expert and data-driven approaches, a table with the estimation in tons of the selected urban crop productivity was carried out for each one of the three scenarios.

Crop	Avg. Harvested yield for urban farmers (kg/m2)
Lettuce	1.45
Chard	5.44
Chickpeas	1.32
Aloe vera	0.12
Cabbage	2.72

Table 11. The average harvested yield of dominant vegetables types grown using urban agricultural

World Health Organization(WHO) and the FAO recommend a daily minimum intake of 400 g of fruits and vegetables per day for the prevention of chronic diseases (WHO/FAO, 2003). Based on this value, the possible number of people benefited was estimated with the values of the urban crop that produced the maximum amount of production by the implementation of UA practices in public open spaces and the population of the urban area for the city of Bogotá (PopulationStat, 2020).

4. RESULTS AND DISCUSSION

While this thesis aimed to implement a land suitability analysis to identify potential available areas for urban agriculture practices in public open spaces using an expert knowledge approach and a data-driven approach, it also aimed to compare the results obtained from both approaches. In this context, the results of the procedure carried out in the expert knowledge approach using the MCDM-AHP method are shown in section 4.1, including a spatial sensitivity analysis in section 4.2 developed to evaluate the robustness of the model implemented. The performance of the machine learning models implemented in the data-driven approach and the MCDM-AHP model are display in section 4.3 and a comparison based on the relevant variables in both approaches is described in section 4.4. The land suitability maps for UA obtained from both approaches and a combined approach are shown in section 4.5. Finally, the estimation of public open spaces and crop productivity is presented in section 4.6 and section 4.7 respectively.

The methodology was implemented using scripts written in Python and R languages. Readers interested in the application of these procedures can visit the following link in GitHub.

4.1 STANDARDIZED CRITERIA AND PAIRWISE COMPARISON MATRIX

The variables classified in the land suitability classes defined for UA (Table 6), according to the values established by the experts (Table 8) and the five main criteria obtained by composed maps of the variables are illustrated in Figure 22. The results of pairwise comparisons assessment of the five main criteria by experts using Saaty's s weighting scale (Table 1) and their corresponding estimated weight are presented in Table 12. The Consistency Ratio (CR) estimated for the comparison matrix was 0.081, which is within the accepted interval of consistency (< 10%), indicating that the relative weights were properly chosen in this land suitability analysis model.

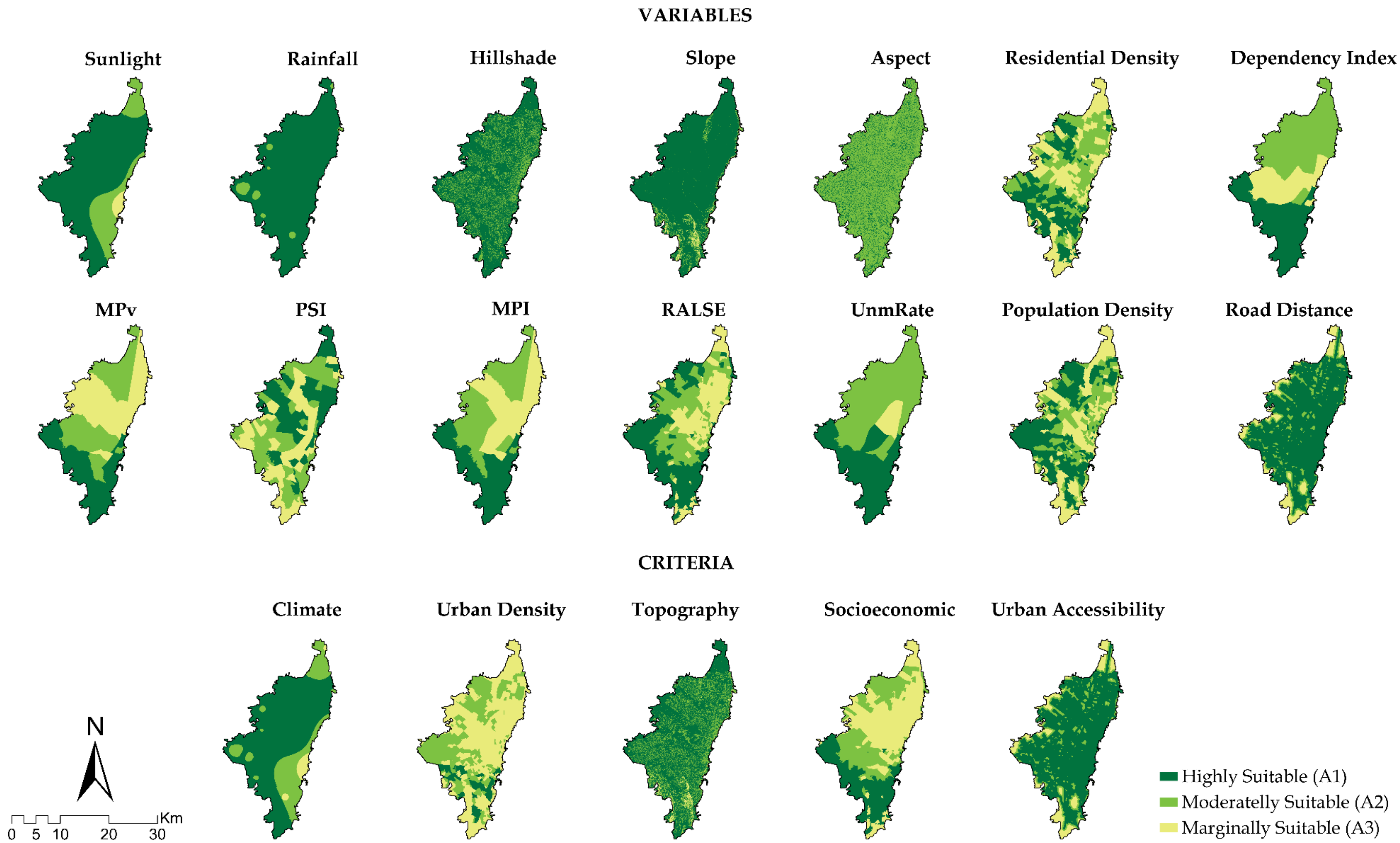

Figure 22. Variables and criteria classified in land suitability classes for UA

CRITERIA COMPARISON MATRIX											
	Socioeconomic	**Topography**	**Urban Density**	**Climate Conditions**	**Urban Accessibility**	**Normalized Matrix**					**Weights**
Socioeconomic	1	3	5	3	5	0.48	0.62	0.52	0.29	0.26	**0.4357**
Topography	1/3	1	3	3	7	0.16	0.21	0.31	0.29	0.37	**0.2677**
Urban Density	1/5	1/3	1	3	3	0.10	0.07	0.10	0.29	0.16	**0.1435**
Climate Conditions	1/3	1/3	1/3	1	3	0.16	0.07	0.03	0.10	0.16	**0.1039**
Urban Accessibility	1/5	1/7	1/3	1/3	1	0.10	0.03	0.03	0.03	0.05	**0.0492**

Table 12. Comparison matrix and estimated weights for the five main criteria selected

4.2 SENSITIVITY ANALYSIS

One-at-a-time (OAT) sensitivity analysis method was implemented to assess the uncertainties in the criteria weights obtained from the expert knowledge approach and determine the robustness of the results. Furthermore, this method allowed to identify which were the most sensitivity criteria and suitability classes in this research. Figure 23 summarizes the result of the sensitivity analysis implemented.

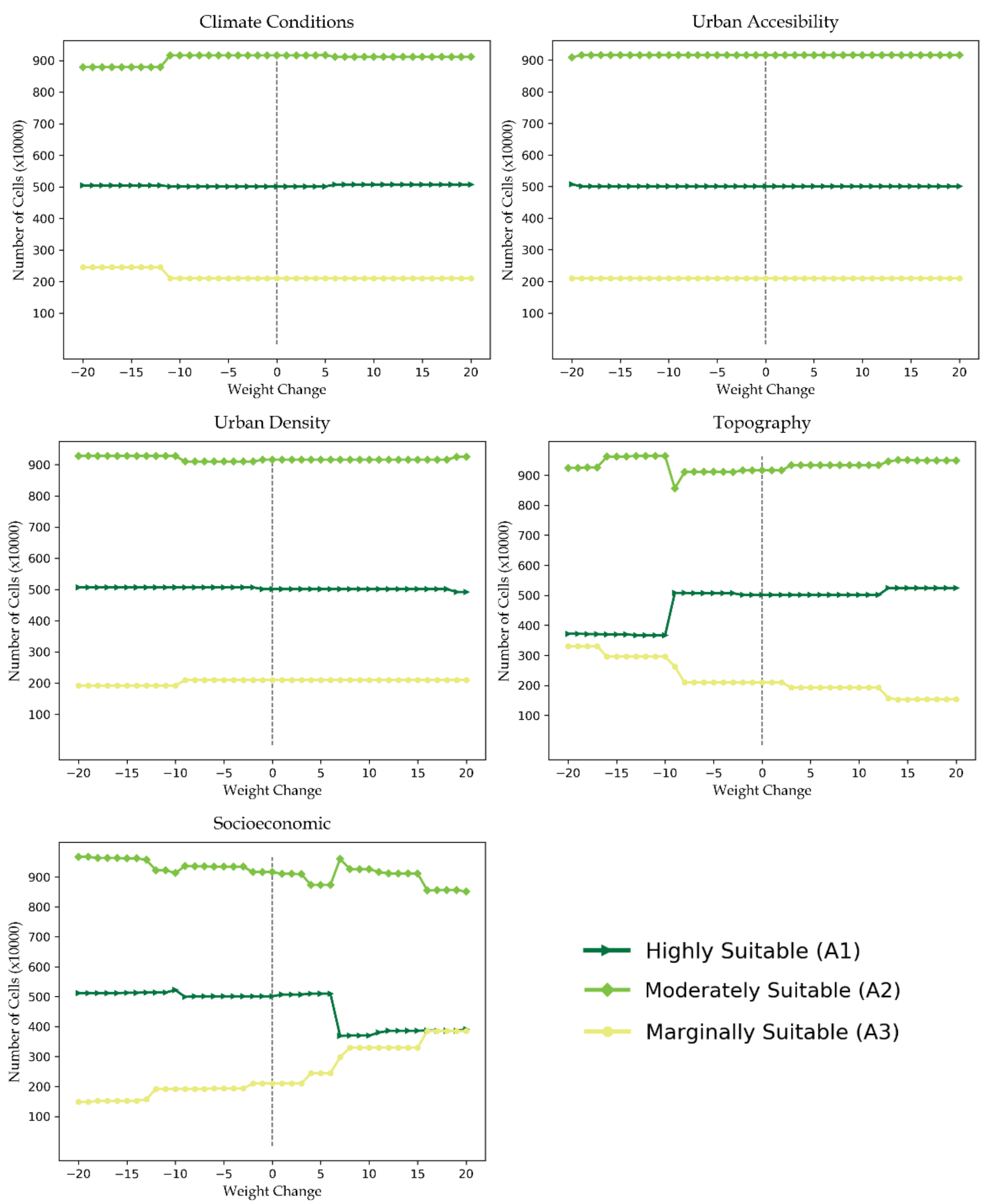

Figure 23. Variations in the suitability classes due to weight changes for each criterion

Although the different variations in the weights of climate conditions, urban density, and urban accessibility criteria, cell values for the suitability classes (A1, A2, and A3) remained relatively stable, being urban density and urban accessibility the lowest sensitivity among all criterion. Furthermore, this indicates that suitability categories have an independent behavior of the variations in the decision weights related to these criteria. On the other hand, the most representative variation among suitability classes could be found in the socioeconomic and topography criteria, where a weight change lower than -10% in the topography criteria modified significantly cell values for the moderately(A2) and marginally (A3) suitability classes, reaching the highest variation in a weight change of -17%. Similarly, a weight change higher than 7% for the socioeconomic criteria shifted considerably cell values for the moderately(A2) and marginally (A3) suitability classes, reaching the highest variation in a weight change of 17%, making this one, the criteria with the highest degree of sensitivity. Figure 24 provides insights about the spatial variation and patterns of sensitivity during the weight variations, helping to confirmed that socioeconomic and topography are the criterion with the highest degree of sensitivity and responsive to changes, having a considerable impact on the resulting maps, modifying the spatial variability on the evaluation results especially between A2 and A3 suitability classes. This might be influenced because of the relative importance of these criteria in the AHP method since both criteria represent the 70% relevance in the expert knowledge approach (Table 12).

Although the sensitivity of these criteria, the results derived from the expert approach (baseline weight criteria) could be used to explore potential locations for UA practices, because variables present in the topography criteria (slope, aspect, and hillshade) remain relatively stable over time ruling out possible variations in their weights by the experts. On the other hand, socioeconomic criteria represents more than 40% of relative importance in the analysis, being very unlikely to be modified their established weight by the experts, because an increase would cause a decrease in the relevance of the other criteria and also in the areas that might have potential for UA practices due to the increase of marginally suitability areas. Additionally, experts ruled out a decrease in the weight of the socioeconomic criteria because this one contains the most relevant variables for the development of UA in the study area and its importance should not be disregarded. Nevertheless, this might be considered as a warning to experts and policymakers, a proper weight assignation for these two criteria is relevant for robust suitability analysis and more accurate results. An example of a summary table was created for the criterion socioeconomic with detail information related to the weights variations, the number of cells in each suitability class, and the number and percentage of cells that have shifted classes compared with the base map, could be found in **Appendix B**.

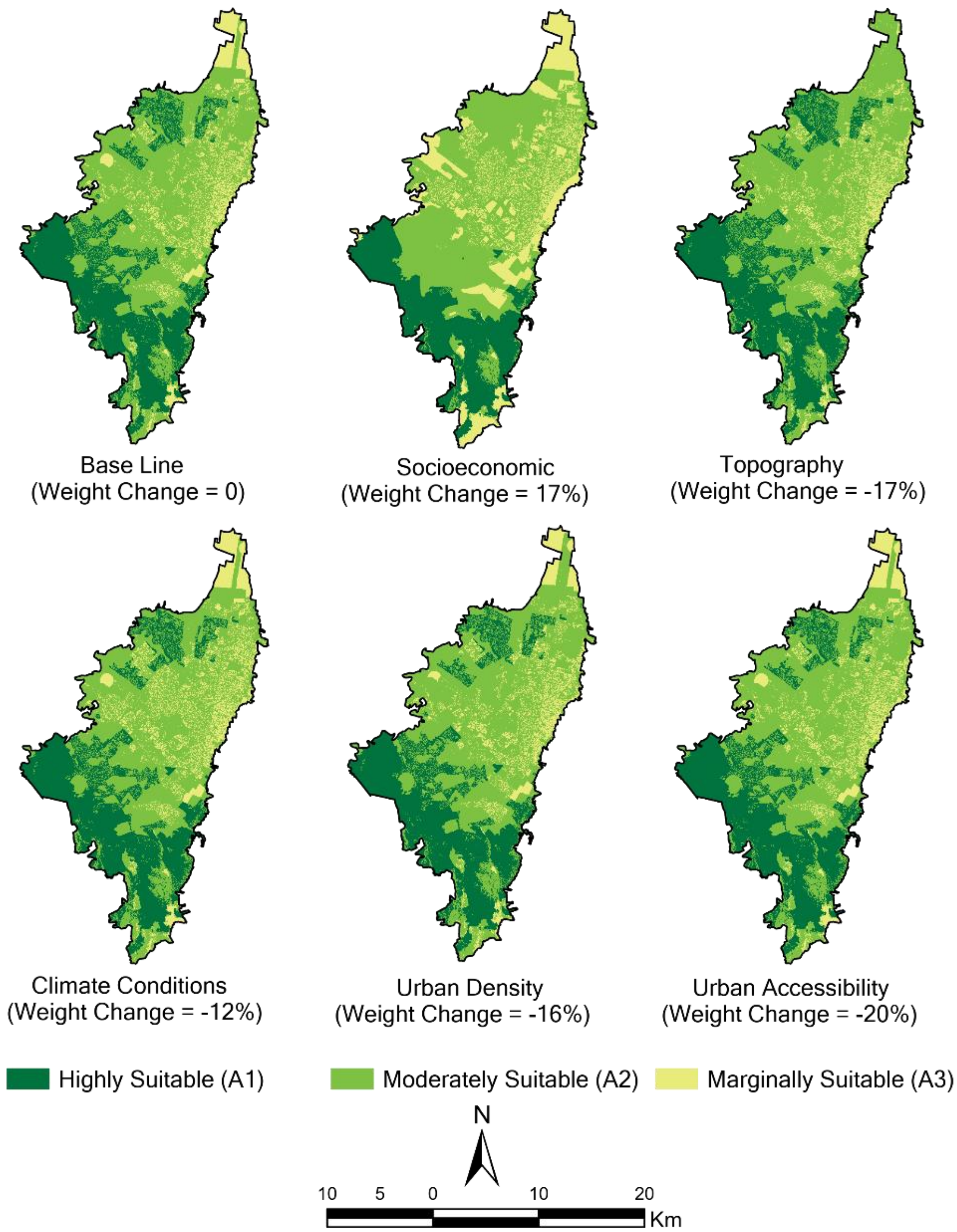

Figure 24. Urban agriculture suitability maps of the maximum changes in evaluation results caused by weights variation in each criterion.

4.3 SELECTION OF THE BEST MODEL

The selection of the best model was obtained by considering the Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) curve, where the AUC values of the models implemented were used to represent the model performance. Figure 25 shows the ROC curves built to statistically compared the models implemented. Among all models, Random Forest (RF) was the model with the best performance with an AUC value of 0.74, followed by SVM (0.71), KNN (0.60) and AHP (0.62). RF was the best model to differentiate between the classes of the *UA class* variable, meaning that there is a 74% probability that this model would be able to distinguish if a location would be considered as a potential candidate for implementation of UA practices. If there is not enough data available, splitting (i.e., training and validation) might have repercussions in the model performance. In this case, not having test data could be more effective (Kuhn & Johnson, 2013). It is not possible to say certainly if the use of all samples in the classifications models, might have affected the performance since the data was not split for training and validation due to the low number of urban orchards in the study area. Instead, the cross-validation method was used for model tuning and validation.

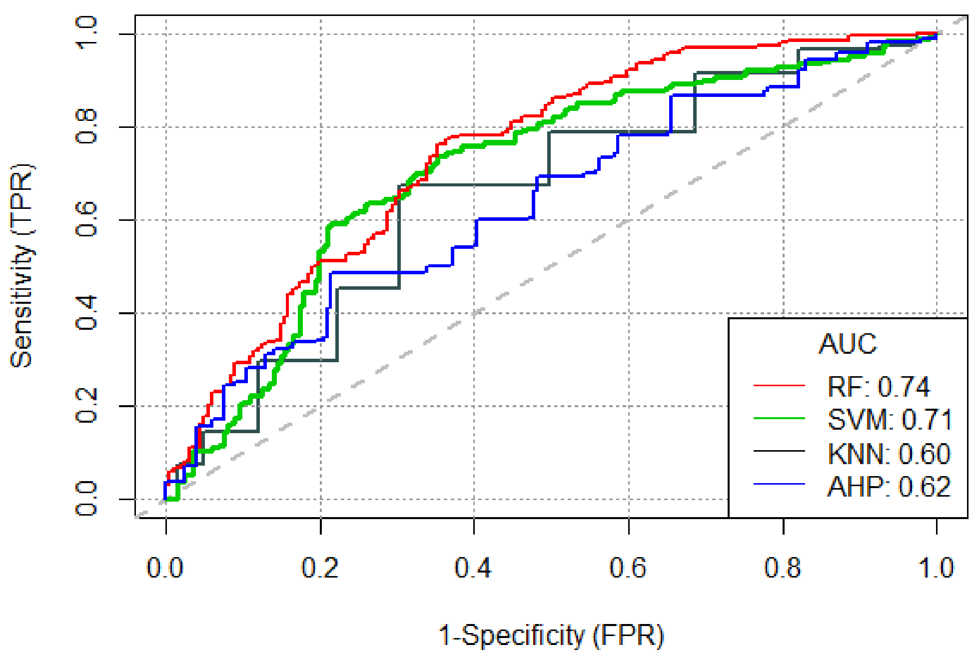

Figure 25. ROC curves and AUC measure performance for Random Forest (RF), Support Vector Machine (SVM), K Nearest Neighbor (KNN) and Analytic Hierarchy Process (AHP)

4.4 MOST RELEVANT VARIABLES

Criteria weights obtained by the MCDM-AHP method were used to represent the variable importance for the expert knowledge approach whereas the random forest model was selected to represent the variable importance for the data-driven approach since it has outperformed the highest accuracy based on AUC results. Figure 26 illustrated the variable importance comparison between both approaches. Understanding that as higher the Importance value higher the predictive power of the predictor variable, it could be seen that Unemployment Rate, Residential Areas with Low Socioeconomic Level (RALSE) and Monetary Poverty are socioeconomic variables that would have a significant contribution to predicting potential sites for UA practices. These indicate that the development of UA is significantly dependent on the socioeconomic conditions of the urban farmers in the study area. Some of the literature has also found an association between population income and UA, pointing out that as a result of poverty, people turn to UA practices for their livelihood and survival (De Zeeuw et al., 2011; FAO, 2011; Orsini et al., 2013). Urban demographics variables also play an important role in the prediction of potential areas for UA practices, which could be explained due to the lack of land. Land availability represents the main constraint for the development of UA in cities (Badami & Ramankutty, 2015). Therefore, availability spaces in parks, green areas, and squares would be highly considered for the development of UA, explaining the important participation of the variable public space indicator in the results. Additionally, the limited availability of land results in intensive production in the available areas, which demands the population workforce (Orsini et al., 2013). Consequently, variables that help to identify population clusters such as residential and population density would be highly relevant for modeling. Based on the results explained above is possible to answer the second research question formulated in this thesis, concluding that the most relevant variables in both approaches are the ones related to the social and economic component that UA is an activity mainly encouraged by social reasons conditioned by economic factors Comparing the results obtained it can be noted that both approaches agreed that socioeconomic and urban density variables are the most relevant to identify potential areas for the development of UA practices, concluding that UA in the study area is an activity mainly encouraged by social reasons conditioned by economic factors.

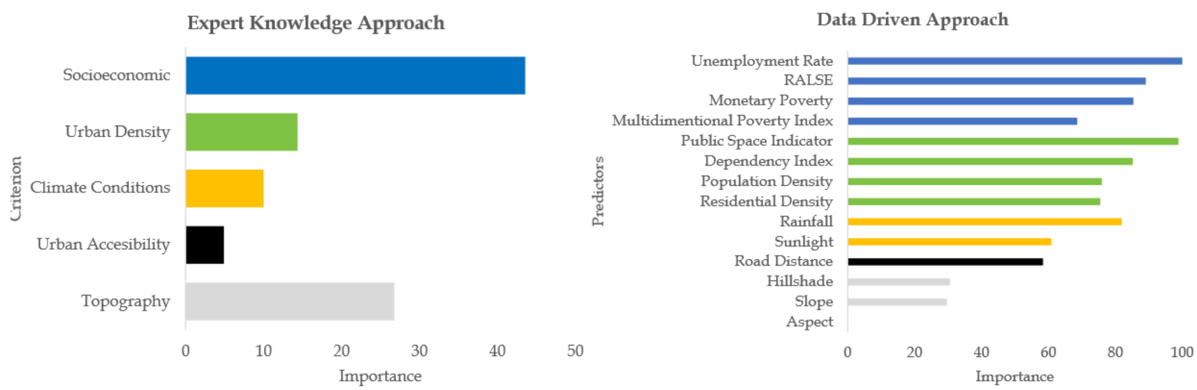

Figure 26. Scaled criteria and predictors variable importance for expert knowledge and data-driven approach

4.5 LAND SUITABILITY MAPS FOR URBAN AGRICULTURE

The random forest model was used to generate the suitability map for UA that represents the data-driven approach since it has outperformed the highest accuracy based on AUC values. Suitability maps for UA obtained with the other classification models can be found in **Appendix C**. The land suitability analysis should be developed in such a way that local needs and conditions are properly reflected in the final decisions (Prakash TN, 2003). The land suitability maps for UA generated for the expert knowledge and data-driven approaches are presented in Figure 27, as well as a combined map of both approaches. The results indicate that most of the highly suitable areas for UA practices are located in the South and Southwestern parts of the city with 21% (8657 ha) based on the expert approach and 18% (7448 ha) based on the data approach. This was an unsurprising result because the variable importance analysis indicated that the socioeconomic and urban demographic variables would be the most relevant and influential in the land suitability results, and this can be evidenced and corroborated with the suitability maps obtained, where the most suitable areas for UA correspond with locations of high concentration of population and dwelling units. Besides, the majority of the socioeconomic variables have their highest values in these areas, whereby it is possible to say that the results of the land suitability analysis meet the expected purpose of poverty alleviation, by improving the food and nutritional security through UA activities because the highly suitable areas for UA practices are located in the part of the city where there is the largest number of people with low incomes or in conditions of poverty. Additionally, these results contributed to answering the first research question of this study related to identifying the most suitable areas for developing UA practices located within the study area.

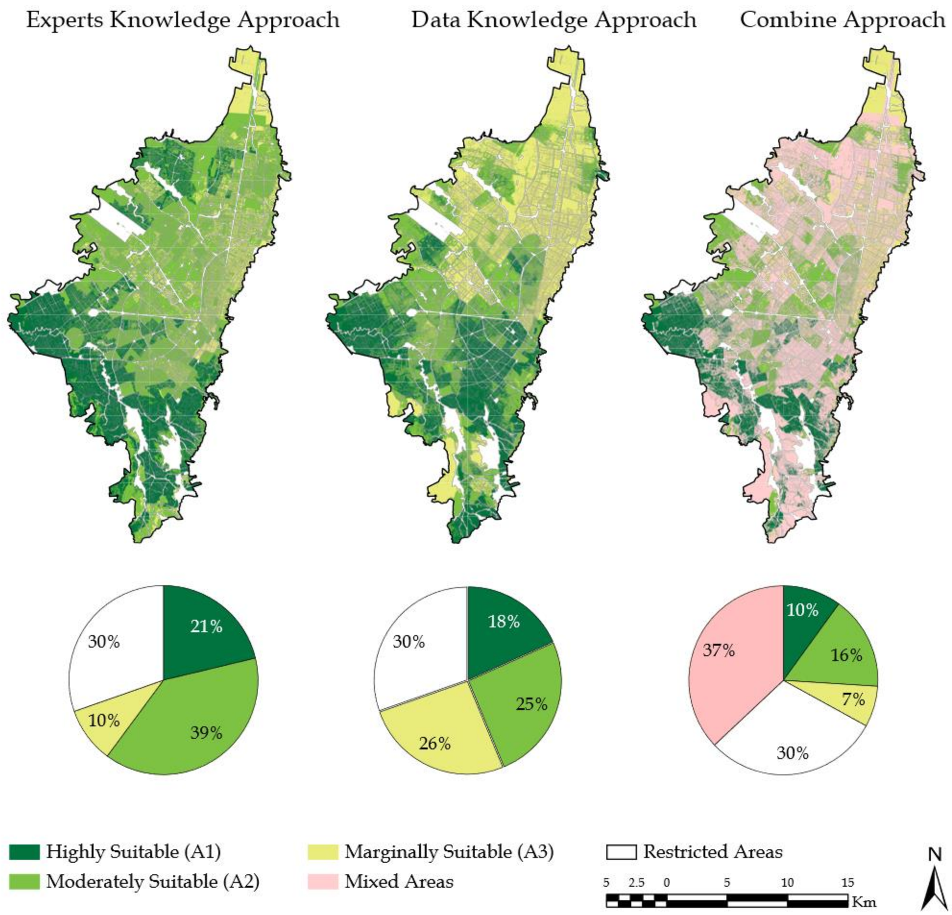

Figure 27. Land Suitability Maps for UA practices

There are significant spatial variations in the suitability classes, especially in the moderately and marginally between both approaches in the center and north part of the city. These variations reflect a considerable decrease in the moderately suitable class of 14% and an increase in the marginally suitable class of 16% from the expert knowledge approach to the data-driven approach. This might be explained mainly by the procedures applied in the data-driven approach to estimate the UA suitability since the model implemented learns and predicts the suitability based on urban orchards data, whereby the high concentration of marginally suitability areas of this approach corresponds with the absence data of urban orchards in the study area.

A compose scenario of common suitability areas in both approaches was created. In this scenario, the highly suitable land for UA represents 10% (4197 ha) of the study area, which means that this land was classified as high suitability either in the expert knowledge and data-driven approach. The pink area in the pie chart (Figure 27), corresponds to changes of suitability classes in the cell values between the two approaches, for example, cell values that were classified as marginally suitable in the expert approach that shifted to moderately or highly suitable in the data-driven approach. Results from this combined approach might be useful to implement policies that prioritize the development of UA in the common suitability areas. Despite the most statistically accurate UA suitability map was obtained by the data-driven approach, the final decision about which approach implements in urban policies oriented to the development of UA practices with poverty alleviation purposes depends on the participation of decision-makers, urban planners, urban farmers, and stakeholders. In the end, negotiating with the landowners of the available areas within the city could be more complicated than identifying potential optimal sites. Nevertheless, this thesis serves as a guide and support for this decision by providing insights into the possible advantages and disadvantages that can be incurred with each approach. For example, if the aim is an implementation that attempts to automate processes and saving time without the involvement of soil scientist experts, a land suitability analysis based on machine learning methods would be recommended (Senagi et al., 2017). On the other hand, if an exhaustive analysis involving all stakeholders without any restrictions is required, a land suitability analysis based on multi-criteria methods would be recommended.

4.6 SUITABLE AREAS FOR URBAN AGRICULTURE IN PUBLIC OPEN SPACES

In total 8164 public open spaces were selected within the study area, of which 86% (Figure 28) are distributed between the public categories of *Parks* and *Educational buildings* (58% and 28% respectively). Statistics of the overlapping process between the land suitability maps for UA and the public open spaces are shown in Figure 29. The results indicate that most of the UA suitable areas available within these spaces are classified as moderately suitable for UA practices with 63% (2484 ha) based on the expert knowledge approach, 42% (1641.6 ha), based on the data-driven approach and 58% (1160.8 ha) for the combined approach (Figure 29A). This is an understandable outcome since the moderately suitable class is the most representative among the UA suitability maps created. Additionally, most of the highly suitable areas are located in the public categories of *Parks* and *Educational buildings* (Figure 29B), which is mainly explained due to the amount of area that these public categories represented in the study area, especially for *Parks,* because most of their area could be implemented for any other type of activity due to the absence of built areas inside them. *Educational building* was the second public category with the most highly suitable areas among all approaches (178.9 ha, 158.3 ha, and 85.6 ha respectively) and based on the information in Figure 10, where 49% of the urban orchards of the city are developed and managed by schools or academic institutions, places this public category as a potential and optimistic scenario for the development and implementation of UA activities in the study area. Further information related to the number of cells, area, and percentage of the UA suitable areas available in public open spaces can be found in **Appendix D**.

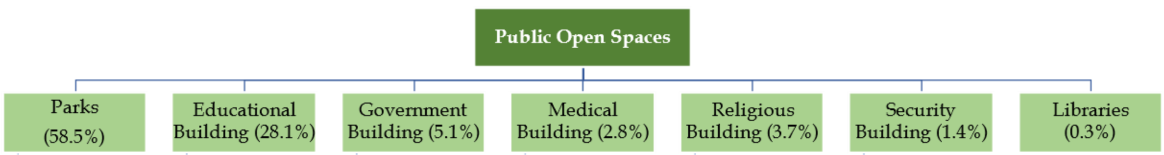

Figure 28. percentage of available areas located in the public categories defined for the open public spaces in the study area

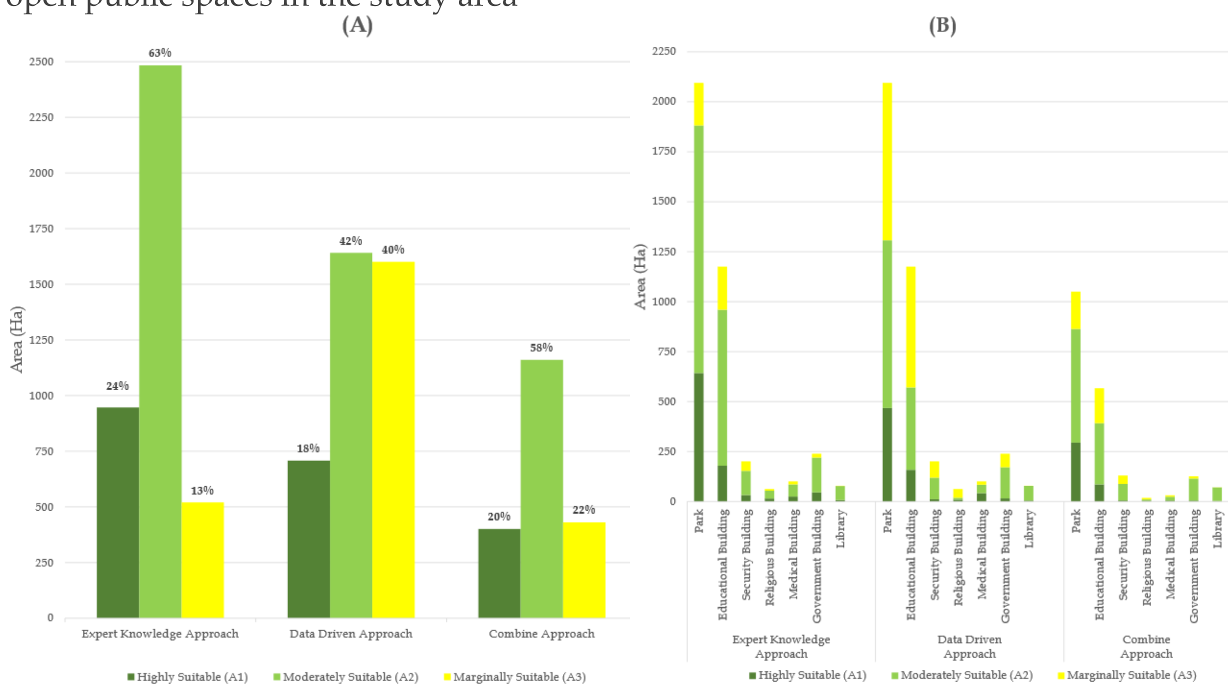

Figure 29. UA suitable areas available in Public Open Spaces

4.7 ESTIMATION OF URBAN CROP PRODUCTION

Based on the information on the average harvested yield of urban crops in Table 11 and the suitable areas available in public open spaces identified in the previous section, an estimate of the production of five urban crops in tons was performed for three possible scenarios. Figure 30A shows the results for the first scenario, where 10% of the total area available in public open spaces and classified as highly suitable for UA activities would be implemented for the development of UA practices.

It could be seen that Chard is the urban crop with the highest productivity among all approaches, with 2575 tons in the expert approach, 1926 tons in the data-driven approach and 1091 tons in the combined approach. These results contributed to answering the third research question of this study related to the vegetable production of public open spaces within the study area. Additionally, the maximum production in tons that can be achieved with each of the selected urban crops in the available areas of the public open spaces classified as highly suitable for UA and grouped by public category is shown in Figure 30B, where the public categories of parks and educational buildings are the most representative due to the amount of area available in public open spaces.

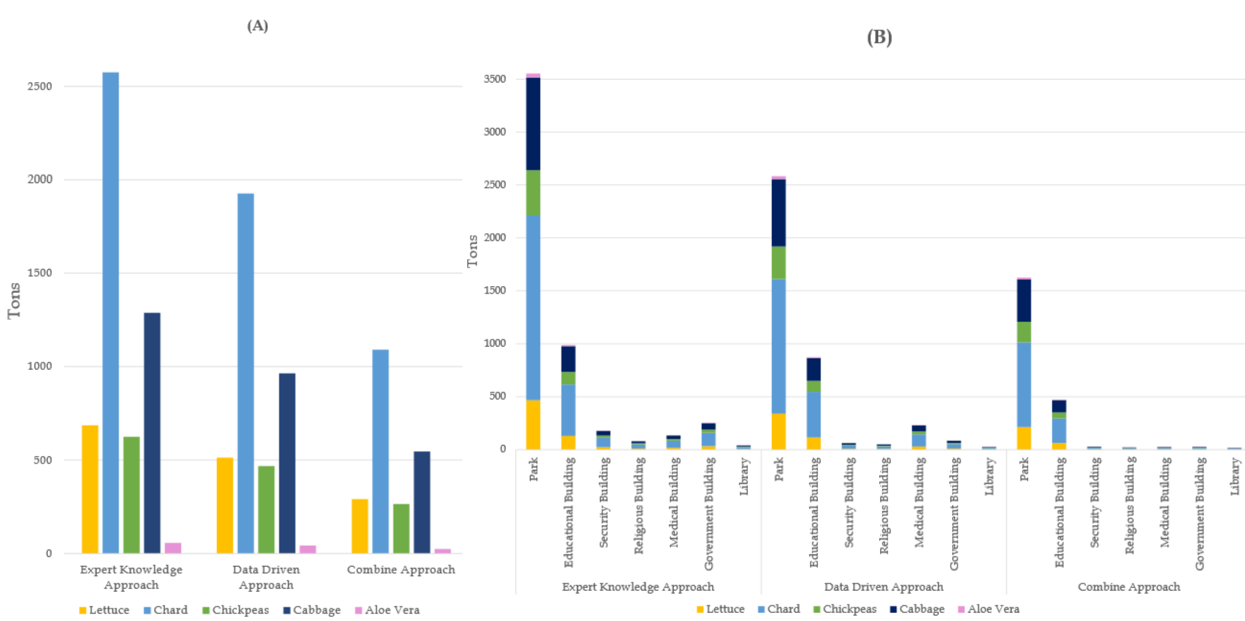

Figure 30. Estimation of urban crops productivity in UA suitable areas available in Public Open Spaces

It is important to understand that the production estimated does not consider the seasonality of the crops mentioned. Therefore, some crops could lead to more production cycles depending on their characteristics. For example, the harvest time of the lettuce varies between 20 – 65 days, whereas for chard and cabbage is between 3 – 5 months. Moreover, in most of the cases, urban orchards sow different types of crops, which could result in the search for the optimization of areas available in public open spaces for the highest possible vegetable production that can be obtained by the combination of different crops, contributing in a greater way to food security in the study area. Finally, using Chard as the urban crop with maximum vegetable production and the less and most optimistic scenarios for the implementation of UA in the study area (10% and 50% respectively) was estimated that between 17.637 and 88.192 people (0.23% and 1.22% respectively) could be benefit of the crop production per year in the study area. For further information related to the number of cells, area, and percentage of the vegetable production estimated in the available areas in public open spaces can be found in **Appendix E**.

5. LIMITATIONS AND FUTURE WORK

In the data-driven approach, the model accuracy is highly dependent on data quality; however, there was the potential of random error from the urban orchards since these are not necessarily being developed in suitable areas for urban agriculture, due to their location is conditioned to random decisions by urban farmers. Furthermore, the limited number of samples where current urban agriculture practices are being developed in the city of Bogotá might induce prediction error in this study. There are also limitations coming from the predicting variables due to the biases of measuring techniques (F. Xu, Ho, Chi, & Wang, 2019). For example, buildings height was estimated using the average height of dwelling units since real high measurements could not be obtained for the study area and subsequently added to a DTM to estimate the hillshade information layer using GIS techniques. A scaling issue coming from the to the socioeconomic variables might also be involved, some variables were rescaled to the parcel level for modeling because their finest level was at the city level. These limitations above may have reduced the accuracy of the modeling results.

Future research should be oriented on data quality and model improvement, including enhancement of data sampling and better selection of predicting variables. For example, feature selection based on machine learning methods could improve the performance and interpretability of the models implemented (Kuhn & Johnson, 2013). This research focused on estimating a production based on the most popular crops grown by urban farmers in the study area. However, a selection of crops that improve food security and nutrition based on their amount of nutrients and calories could be implemented.

Image classification techniques should be implemented to identify the type of land in the public open spaces because the costs of implementing urban agriculture activities vary according to the type of soil. For example, locations with no natural soil should involve in additional costs for the purchase and transport of containers to store the nutrients required for the crops. In addition, an extended analysis should be undertaken on the potential of private open spaces to encourage landowners to implement practices o urban agriculture in their properties such as potential tax deduction, credit facility or monetary incentives.

6. CONCLUSIONS

The thesis aimed at answering three research questions that in a general-purpose seek to identify potential available areas for urban agriculture practices within public open spaces. Two approaches were proposed in the methodology to estimate the land suitability of the city of Bogotá for urban agriculture activities. The first approach implements a subjective method based on a Multicriteria Decision Making Analysis (MCDM) and the second approach applied an objective assessment derived from historical data using machine learning techniques. The result of applying this methodology depicts that for the first approach 21% of the study area has highly suitable land for urban agriculture activities, 39% moderately suitable and 10% marginally suitable; for the second approach, 18% of the study area has highly suitable land for urban agriculture activities, 25% moderately suitable and 26% marginally suitable. Both approaches coincided that the highly suitable areas for urban agriculture practices are located in the South and Southwest side of the city. The resulting suitability maps and statistics lead to answer the first research question related to the location of the most suitable areas for the development of urban agriculture activities in the study area.

To answer the second research question, this thesis proposed a comparison analysis between both approaches. In this analysis, a statistical comparison based on the Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) curve was used as a measure to compare the performance of the methods applied. The results showed that random forest (RF) algorithm used in the second approach had the highest accuracy with 0.74 based on the AUC values. Additionally, a visual comparison was carried out based on the relevant variables in each approach indicating that socioeconomic and demographic variables are the most relevant criteria for urban agriculture.

The last stage of the designed methodology is aimed at answering the last research question. To do so, GIS techniques were implemented to identify the available areas within public open spaces that overlay with the suitability land for urban agriculture obtained from both approaches. Three possible scenarios for the development of urban agriculture practices in these areas were evaluated. The first scenario states that 10% of the areas available in public open spaces would be used for future urban agriculture activities, the second and third scenarios vary only in the percentage of area implemented with 30% and 50% respectively. Using the average harvested yield for five of the most popular crops grown by urban agriculture, vegetable productivity was estimated for the three scenarios established and the resulting statistics lead to answer the third research question related to the vegetable production of public open spaces within the study area.

This thesis assessed what could be the potential of open public spaces in the possible implementation of urban agriculture practices in the city of Bogota, providing useful information for urban planning policies and decision-makers geared to achieve multifunctional and sustainable land use for current public open spaces. An extent analysis should be undertaken on the potential of private spaces to encourage landowners to implement practices of urban agriculture in their properties such as potential tax deduction, credit facility or monetary incentives. Moreover, the results may become an input for local and governmental entities as support for the inclusion of spaces for urban agriculture within urban planning policies oriented in improve the food security and nutrition in the city, generating opportunities for the establishment of local economies that contribute to the reduction of unemployment and urban poverty.

BIBLIOGRAPHIC REFERENCES

Abilash, R. (2018). Applying Random Forest (Classification) — Machine Learning Algorithm from Scratch with Real Datasets. Retrieved February 20, 2020, from https://medium.com/@ar.ingenious/applying-random-forest-classification-machine-learning-algorithm-from-scratch-with-real-24ff198a1c57

Akinci, H., Özalp, A. Y., & Turgut, B. (2013). Agricultural land use suitability analysis using GIS and AHP technique. *Computers and Electronics in Agriculture*, *97*, 71–82. https://doi.org/10.1016/j.compag.2013.07.006

Aldababseh, A., Temimi, M., Maghelal, P., Branch, O., & Wulfmeyer, V. (2018). Multi-criteria evaluation of irrigated agriculture suitability to achieve food security in an arid environment. *Sustainability (Switzerland)*, *10*(3). https://doi.org/10.3390/su10030803

Badami, M. G., & Ramankutty, N. (2015). Urban agriculture and food security: A critique based on an assessment of urban land constraints. *Global Food Security*, *4*, 8–15. https://doi.org/10.1016/j.gfs.2014.10.003

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, *3*(2), 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Bogota Botanical Garden. (2007). *Agricultura Urbana: Cartilla técnica*. Retrieved from http://www.jbb.gov.co/documentos/tecnica/2018/cartilla_tecnica_agricultura_urbana.pdf

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cantor, K. M. (2010). Agricultura urbana: Elementos valorativos sobre su sostenibilidad. *Cuadernos de Desarrollo Rural*, *7*(65), 59–84. Retrieved from http://www.scielo.org.co/pdf/cudr/v7n65/v7n65a04.pdf

Castelli, M., Vanneschi, L., & Largo, Á. R. (2019). Supervised Learning: Classification. In *Encyclopedia of Bioinformatics and Computational Biology* (pp. 342–349). https://doi.org/10.1016/B978-0-12-809633-8.20332-4

Chen, Y., Yu, J., & Khan, S. (2010). Spatial sensitivity analysis of multi-criteria weights in GIS-based land suitability evaluation. *Environmental Modelling and Software*, *25*(12), 1582–1591. https://doi.org/10.1016/j.envsoft.2010.06.001

Chen, Yun, Yu, J., & Khan, S. (2013). The spatial framework for weight sensitivity analysis in AHP-based multi-criteria decision making. *Environmental Modelling and Software*, *48*, 129–140. https://doi.org/10.1016/j.envsoft.2013.06.010

Chiranjit, C. C. (2015). *Data Classification* (Chapman and Hall/CRC, Ed.). Retrieved from https://doi.org/10.1201/b17320

Chivasa, W., Mutanga, O., & Biradar, C. (2019). Mapping land suitability for maize (Zea mays L.) production using GIS and AHP technique in Zimbabwe. *South African Journal of Geomatics*, *8*(2), 265–281. https://doi.org/10.4314/sajg.v8i2.11

Crosetto, M., Tarantola, S., & Saltelli, A. (2000). Sensitivity and uncertainty analysis in spatial modelling based on GIS. *Agriculture, Ecosystems and Environment*. https://doi.org/10.1016/S0167-8809(00)00169-9

de Figueiredo, M., Cordella, C. B. Y., Jouan-Rimbaud Bouveresse, D., Archer, X., Bégué, J. M., & Rutledge, D. N. (2018). A variable selection method for multiclass classification problems using two-class ROC analysis. *Chemometrics and Intelligent Laboratory Systems*, *177*, 35–46. https://doi.org/10.1016/j.chemolab.2018.04.005

De Zeeuw, H., Van Veenhuizen, R., & Dubbeling, M. (2011). The role of urban agriculture in building resilient cities in developing countries. *Journal of Agricultural Science*, *149*(S1), 153–163. https://doi.org/10.1017/S0021859610001279

Dongus, D. S., & A. (2009). The use of GIS, GPS and aerial imagery for mapping urban agricultural activities on open space in cities - Geospatial World. Retrieved January 24, 2020, from https://www.geospatialworld.net/article/the-use-of-gis-gps-and-aerial-imagery-for-mapping-urban-agricultural-activities-on-open-space-in-cities/

Ermini, P. V., Delprino, M. R., & Giobellina, B. (2017). Mapeo de la agricultura urbana y periurbana en el área metropolitana Santa Rosa-Toay: Aproximaciones metodológicas para la lectura territorial. *Revista de Investigaciones Agropecuarias*, *43*(3), 280–290.

ESRI. (2020). ArcGIS Pro | 2D and 3D GIS Mapping Software. Retrieved January 10, 2020, from ArcGIS Pro website: https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview

Fadhil, D. N., & Moeckel, R. (2018). *A GIS-based Analysis for Selecting Ground Infrastructure Locations for Urban Air Mobility*. (May). Retrieved from http://www.msm.bgu.tum.de/fileadmin/w00bvh/www/masterThesis/fadhil_2018.pdf

FAO. (2007). *Land Evaluation: Towards a Revised Framework*. Retrieved from http://www.fao.org/nr/lman/docs/lman_070601_en.pdf

FAO. (2011). *The Place of Urban and Peri-Urban Agriculture (UPA) in National Food Security Programmes*. Retrieved from http://www.fao.org/3/i2177e/i2177e00.pdf

Feizizadeh, B., & Blaschke, T. (2013). Land suitability analysis for Tabriz County, Iran: A multi-criteria evaluation approach using GIS. *Journal of Environmental Planning and Management*, *56*(1), 1–23. https://doi.org/10.1080/09640568.2011.646964

Foster, A. D., & Rosenzweig, M. R. (2007). Chapter 47 Economic Development and the Decline of Agricultural Employment. *Handbook of Development Economics*, *4*(07), 3051–3083. https://doi.org/10.1016/S1573-4471(07)04047-8

Gahukar, G. (2018). Classification Algorithms in Machine Learning. Retrieved February 21, 2020, from https://medium.com/datadriveninvestor/classification-algorithms-in-machine-learning-85c0ab65ff4

Goldscheider, C. (2019). *Urban migrants in developing nations: patterns and problems of adjustment*. Routledge.

Hemakumara, G. (2015). Gis Based Analysis For Suitability Location Finding In The Residential Development Areas Of Greater Matara Region. *International Journal of Scientific & Technology Research*, *4*(8), 96–105.

Heumann, B. W., Walsh, S. J., & McDaniel, P. M. (2011). Assessing the application of a geographic presence-only model for land suitability mapping. *Ecological Informatics*, *6*(5), 257–269. https://doi.org/10.1016/j.ecoinf.2011.04.004

Hijmans, R. J. (2020). *Package "raster" Type Package Title Geographic Data Analysis and Modeling*. Retrieved from https://cran.r-project.org/package=raster

Ingersoll, R. (2012). Urban agriculture. *Lotus International*, *149*, 105–117.

Jafari, S., & Zaredar, N. (2010). Land Suitability Analysis using Multi Attribute Decision Making Approach. *International Journal of Environmental Science and Development*, *1*(5), 441–445. https://doi.org/10.7763/ijesd.2010.v1.85

Kuhn, M. (2019). The caret Package. Retrieved January 11, 2020, from http://topepo.github.io/caret/index.html

Kuhn, M. (2020). A Short Introduction to the caret Package. Retrieved January 10, 2020, from https://cran.r-project.org/web/packages/caret/vignettes/caret.html

Kuhn, M., & Johnson, K. (2013a). Applied predictive modeling. In *Applied Predictive Modeling*. https://doi.org/10.1007/978-1-4614-6849-3

Kuhn, M., & Johnson, K. (2013b). Applied predictive modeling. In *Applied Predictive Modeling*. https://doi.org/10.1007/978-1-4614-6849-3

La Rosa, D., Barbarossa, L., Privitera, R., & Martinico, F. (2014). Agriculture and the city: A method for sustainable planning of new forms of agriculture in urban contexts. *Land Use Policy*, *41*, 290–303. https://doi.org/10.1016/j.landusepol.2014.06.014

Lee-Smith, D. (2010). Cities Feeding People: An Update on Urban Agriculture in Equatorial Africa. *Environment and Urbanization - Environ Urban*, *22*, 483–499.

https://doi.org/10.1177/0956247810377383

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, *18*(8), 2674.

Liu, Y., Lv, X., Qin, X., Guo, H., Yu, Y., Wang, J., & Mao, G. (2007). An integrated GIS-based analysis system for land-use management of lake areas in urban fringe. *Landscape and Urban Planning*, *82*(4), 233–246. https://doi.org/10.1016/j.landurbplan.2007.02.012

Malczewski, J. (2006). GIS-based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science*, *20*(7), 703–726. https://doi.org/10.1080/13658810600661508

Matplotlib.org. (2020). Matplotlib: Python plotting — Matplotlib 3.1.3 documentation. Retrieved February 16, 2020, from https://matplotlib.org/

McClintock, N., Cooper, J., & Khandeshi, S. (2013). Assessing the potential contribution of vacant land to urban vegetable production and consumption in Oakland, California. *Landscape and Urban Planning*, *111*(1), 46–58. https://doi.org/10.1016/j.landurbplan.2012.12.009

McEniry, M., Samper-Ternent, R., & Cano-Gutierrez, C. (2019). Displacement due to armed conflict and violence in childhood and adulthood and its effects on older adult health: The case of the middle-income country of Colombia. *SSM - Population Health*, *7*. https://doi.org/10.1016/j.ssmph.2019.100369

Mendas, A., & Delali, A. (2012). Integration of MultiCriteria Decision Analysis in GIS to develop land suitability for agriculture: Application to durum wheat cultivation in the region of Mleta in Algeria. *Computers and Electronics in Agriculture*. https://doi.org/10.1016/j.compag.2012.02.003

Mokarram, M., Hamzeh, S., Aminzadeh, F., & Zarei, A. R. (2015). Using machine learning for land suitability classification. *West African Journal of Applied Ecology*, *23*(1), 63–73.

Montgomery, B., & Schmidt, M. (2015). *Expanding and comparing GIS-based multi - criteria decision making methods : A soft computing logic for agricultural land suitability evaluation by Summer 2015 Approval*.

Orsini, F., Kahane, R., Nono-Womdim, R., & Gianquinto, G. (2013a). Urban agriculture in the developing world: A review. *Agronomy for Sustainable Development*, *33*(4), 695–720. https://doi.org/10.1007/s13593-013-0143-z

Orsini, F., Kahane, R., Nono-Womdim, R., & Gianquinto, G. (2013b). Urban agriculture in the developing world: a review Urban agriculture in the developing world: a review. Agronomy for Sustainable Development, Springer Verlag/EDP Sci-ences/ Urban agriculture in the developing world: a review. *INRA*, *33*(4).

https://doi.org/10.1007/s13593-013-0143-zï

Pantoulas, D., & McCoy, J. (2019). Venezuela: an unstable equilibrium. *Revista de Ciencia Política*, *39*, 391–408.

Parthiban, C., & Krishnan, M. B. (2016). Expert System for Land Suitability Evaluation using Data mining's Classification Techniques: a Comparative Study. *International Journal of Computer Trends and Technology*, *33*(2), 87–92. https://doi.org/10.14445/22312803/ijctt-v33p119

Pavan, M. (2009). Multi-Criteria Decision Making - an overview | ScienceDirect Topics. Retrieved January 23, 2020, from https://www.sciencedirect.com/topics/mathematics/multi-criteria-decision-making

Pontius, R. G., & Parmentier, B. (2014). Recommendations for using the relative operating characteristic (ROC). *Landscape Ecology*, *29*(3), 367–382. https://doi.org/10.1007/s10980-013-9984-8

PopulationStat. (2020). Bogota, Colombia Population (2020) - Population Stat. Retrieved February 23, 2020, from https://populationstat.com/colombia/bogota

Prakash TN. (2003). *Land Suitability Analysis for Agricultural Crops: A Fuzzy Multicriteria Decision Making Approach*.

Puntsag, G., Kristjánsdóttir, S., & Ingólfsdóttir, B. D. (2014). *Land Suitability Analysis for Urban and Agricultural Land Using Gis: Case Study in Hvita To Hvita, Iceland*. Retrieved from http://www.unulrt.is/static/fellows/document/puntsag2014.pdf

Python.org. (2020). About Python™ | Python.org. Retrieved January 10, 2020, from https://www.python.org/about/

Saaty, R. W. (1987). The analytic hierarchy process-what it is and how it is used. *Mathematical Modelling*, *9*(3–5), 161–176. https://doi.org/10.1016/0270-0255(87)90473-8

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, *15*(3), 234–281. https://doi.org/10.1016/0022-2496(77)90033-5

Sarmadian, F., Keshavarzi, A., Rooien, A., Zahedi, G., Javadikia, H., & Iqbal, M. (2014). *Support Vector Machines Based-Modeling of Land Suitability Analysis for Rainfed Agriculture*. *2*(4), 165–171. https://doi.org/10.12691/jgg-2-4-4

Senagi, K., Jouandeau, N., & Kamoni, P. (2017). Using parallel random forest classifier in predicting land suitability for crop production. *Journal of Agricultural Informatics*, *8*(3), 23–32. https://doi.org/10.17700/jai.2017.8.3.390

Setiawan, A., Sediyono, E., & A. L. Moekoe, D. (2014). Application of AHP Method in Determining Priorities of Conversion of Unusedland to Food Land in Minahasa Tenggara. *International Journal of Computer Applications*, *89*(8), 37–44. https://doi.org/10.5120/15526-4433

Shalev-Shwartz, S., & Ben-David, S. (2013). Understanding machine learning: From theory to algorithms. In *Understanding Machine Learning: From Theory to Algorithms* (Vol. 9781107057). https://doi.org/10.1017/CBO9781107298019

Singha, C., & Swain, K. C. (2016). *Land suitability evaluation criteria for agricultural crop selection: A review*. https://doi.org/10.18805/ar.v37i2.10737

Spataru, A., Faggian, R., & Sposito, V. (2018). *Agricultural Land Suitability Analysis of Metropolitan Peri-Urban areas now and into the future-Case study of City of Whittlesea , Melbourne , Australia Agricultural Land Suitability Analysis of Metropolitan Peri-Urban areas now and into the future – Case*. (May), 80–88.

Spatial analysis in ArcGIS Pro—ArcGIS Pro | ArcGIS Desktop. (2018). Retrieved January 10, 2020, from ArcGIS Pro website: https://pro.arcgis.com/en/pro-app/help/analysis/introduction/spatial-analysis-in-arcgis-pro.htm

Thomaier, S., Specht, K., Henckel, D., Dierich, A., Siebert, R., Freisinger, U. B., & Sawicka, M. (2015a). Farming in and on urban buildings: Present practice and specific novelties of zero-acreage farming (ZFarming). *Renewable Agriculture and Food Systems*, *30*(1), 43–54. https://doi.org/10.1017/S1742170514000143

Thomaier, S., Specht, K., Henckel, D., Dierich, A., Siebert, R., Freisinger, U. B., & Sawicka, M. (2015b). Farming in and on urban buildings: Present practice and specific novelties of Zero-Acreage Farming (ZFarming). *Renewable Agriculture and Food Systems*, *30*(1), 43–54. https://doi.org/10.1017/S1742170514000143

Thomas, G. (2014). *An FAO report on urban and peri-urban agriculture in the region*. Retrieved from www.fao.org/publications

Thornton, A., Momoh, J., & Tengbe, P. (2012). Institutional capacity building for urban agriculture research using Participatory GIS in a post-conflict context: a case study of Sierra Leone. *Australasian Review of African Studies*, *33*(1), 165–174. Retrieved from http://search.proquest.com/docview/1373467158?accountid=13042%5Cnhttp://oxfordsfx.hosted.exlibrisgroup.com/oxford?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ:pais&atitle=Institutional+capacity+building+for+urba

Toms, S. (2015). *ArcPy and ArcGIS – Geospatial Analysis with Python*.

Toole, M. (2019). Health in humanitarian crises. *The Health of Refugees: Public Health Perspectives from Crisis to Settlement*.

Uy, P. D., & Nakagoshi, N. (2008a). Application of land suitability analysis and landscape ecology to urban greenspace planning in Hanoi, Vietnam. *Urban Forestry and Urban Greening*, *7*(1), 25–40. https://doi.org/10.1016/j.ufug.2007.09.002

Uy, P. D., & Nakagoshi, N. (2008b). Application of land suitability analysis and landscape ecology to urban greenspace planning in Hanoi, Vietnam. *Urban Forestry and Urban Greening*, *7*(1), 25–40. https://doi.org/10.1016/j.ufug.2007.09.002

Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science and Engineering*, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37

Van Veenhuizen, R., & Danso, G. (2007). *Profitability and sustainability of urban and periurban agriculture* (Vol. 19). Food & Agriculture Org.

Vargas, R. V. (2010). *Using the Analytic Hierarchy Process (Ahp) to Select and Prioritize Projects in a Portfolio*.

WHO/FAO. (2003). *Diet, Nutrition and the Prevention of Chronic Diseases*.

Wickham, H. (2020). Create Elegant Data Visualisations Using the Grammar of Graphics ggplot2. Retrieved January 10, 2020, from https://ggplot2.tidyverse.org/

Wong, T. T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, *48*(9), 2839–2846. https://doi.org/10.1016/j.patcog.2015.03.009

Xu, E., & Zhang, H. (2013). Spatially-explicit sensitivity analysis for land suitability evaluation. *Applied Geography*, *45*, 1–9. https://doi.org/10.1016/j.apgeog.2013.08.005

Xu, F., Ho, H. C., Chi, G., & Wang, Z. (2019). Abandoned rural residential land: Using machine learning techniques to identify rural residential land vulnerable to be abandoned in mountainous areas. *Habitat International*, *84*(October 2018), 43–56. https://doi.org/10.1016/j.habitatint.2018.12.006

APPENDIX A: CRITERIA DECISION TREES

Roads Distance - Meters(m)	Suitability Class
≤ 200	A1
> 200 - ≤ 500	A2
> 500	A3

Table A 1. Urban accessibility criteria decision tree

Slope Percentage (0 – 100%)	Hillshade (Presence of shadows)	Aspect (Degrees 0 -360)	Suitability Class
≤ 25	No Shadows	North (0–22.5; 337.5–360) South (112.5–247.5)	A1
		West 22.5–112.5 East 247.5–337.5	
	Once or Twice per day	North (0–22.5; 337.5–360) South (112.5–247.5)	A2
		West 22.5–112.5 East 247.5–337.5	
	Three times per day	North (0–22.5; 337.5–360) South (112.5–247.5)	
		West 22.5–112.5 East 247.5–337.5	
> 25 - ≤ 50	No Shadows	North (0–22.5; 337.5–360) South (112.5–247.5)	
		West 22.5–112.5 East 247.5–337.5	
	Once or Twice per day	North (0–22.5; 337.5–360) South (112.5–247.5)	
		West 22.5–112.5 East 247.5–337.5	
	Three times per day	North (0–22.5; 337.5–360) South (112.5–247.5)	
		West 22.5–112.5 East 247.5–337.5	
> 50	No Shadows	North (0–22.5; 337.5–360) South (112.5–247.5)	A3
		West 22.5–112.5 East 247.5–337.5	
	Once or Twice per day	North (0–22.5; 337.5–360) South (112.5–247.5)	A3
		West 22.5–112.5 East 247.5–337.5	
	Three times per day	North (0–22.5; 337.5–360) South (112.5–247.5)	A3
		West 22.5–112.5 East 247.5–337.5	

Table A 2. Topography criteria decision tree

Population Density (People/ha)	Residential Density (Dwelling units / ha)	Dependence Index (31.58 – 50.12)	PSI (People/m²)	Suitability Class	Population Density (People/ha)	Residential Density (Dwelling units / ha)	Dependence Index (31.58 – 50.12)	PSI (People/m²)	Suitability Class	Population Density (People/ha)	Residential Density (Dwelling units / ha)	Dependence Index (31.58 – 50.12)	PSI (People/m²)	Suitability Class
≥ 200	≥ 25	≥ 45	≥ 6	A1	≥ 100 - < 200	≥ 25	≥ 45	≥ 6	A3	< 100	≥ 25	≥ 45	≥ 6	A3
			≥ 3 - < 6					≥ 3 - < 6					≥ 3 - < 6	
			< 3					< 3					< 3	
		≥ 40 - < 45	≥ 6				≥ 40 - < 45	≥ 6				≥ 40 - < 45	≥ 6	
			≥ 3 - < 6					≥ 3 - < 6					≥ 3 - < 6	
			< 3	A2				< 3					< 3	
		< 40	≥ 6				< 40	≥ 6				< 40	≥ 6	
			≥ 3 - < 6					≥ 3 - < 6					≥ 3 - < 6	
			< 3					< 3					< 3	
	≥ 5 - < 25	≥ 45	≥ 6			≥ 5 - < 25	≥ 45	≥ 6			≥ 5 - < 25	≥ 45	≥ 6	
			≥ 3 - < 6					≥ 3 - < 6					≥ 3 - < 6	
			< 3					< 3					< 3	
		≥ 40 - < 45	≥ 6				≥ 40 - < 45	≥ 6				≥ 40 - < 45	≥ 6	
			≥ 3 - < 6					≥ 3 - < 6					≥ 3 - < 6	
			< 3	A3				< 3					< 3	
		< 40	≥ 6				< 40	≥ 6				< 40	≥ 6	
			≥ 3 - < 6					≥ 3 - < 6					≥ 3 - < 6	
			< 3					< 3					< 3	
	< 5	≥ 45	≥ 6			< 5	≥ 45	≥ 6			< 5	≥ 45	≥ 6	
			≥ 3 - < 6					≥ 3 - < 6					≥ 3 - < 6	
			< 3					< 3					< 3	
		≥ 40 - < 45	≥ 6				≥ 40 - < 45	≥ 6				≥ 40 - < 45	≥ 6	
			≥ 3 - < 6					≥ 3 - < 6					≥ 3 - < 6	
			< 3					< 3					< 3	
		< 40	≥ 6				< 40	≥ 6				< 40	≥ 6	
			≥ 3 - < 6					≥ 3 - < 6					≥ 3 - < 6	
			< 3					< 3					< 3	

Table A 3. Urban density criteria decision tree

RALSE Index (0 – 0.6)	Unemployment Rate (4.3 – 13.55)	MPI Index (0.6 – 10.9)	MPv Percentage (3.06 – 33.85)	Suitability Class	RALSE Index (0 – 0.6)	Unemployment Rate (4.3 – 13.55)	MPI Index (0.6 – 10.9)	MPv Percentage (3.06 – 33.85)	Suitability Class	RALSE Index (0 – 0.6)	Unemployment Rate (4.3 – 13.55)	MPI Index (0.6 – 10.9)	MPv Percentage (3.06 – 33.85)	Suitability Class
≥ 0.4	≥ 10	≥ 5	≥ 20	A1	≥ 0.2 - < 0.4	≥ 10	≥ 5	≥ 20	A2	< 0.2	≥ 10	≥ 5	≥ 20	A3
			≥ 10 - < 20					≥ 10 - < 20					≥ 10 - < 20	
			< 10					< 10					< 10	
		≥ 3 - < 5	≥ 20				≥ 3 - < 5	≥ 20				≥ 3 - < 5	≥ 20	
			≥ 10 - < 20					≥ 10 - < 20					≥ 10 - < 20	
			< 10					< 10					< 10	
		< 3	≥ 20				< 3	≥ 20				< 3	≥ 20	
			≥ 10 - < 20					≥ 10 - < 20					≥ 10 - < 20	
			< 10					< 10					< 10	
	≥ 5 - < 10	≥ 5	≥ 20	A2		≥ 5 - < 10	≥ 5	≥ 20			≥ 5 - < 10	≥ 5	≥ 20	
			≥ 10 - < 20					≥ 10 - < 20					≥ 10 - < 20	
			< 10					< 10					< 10	
		≥ 3 - < 5	≥ 20				≥ 3 - < 5	≥ 20				≥ 3 - < 5	≥ 20	
			≥ 10 - < 20					≥ 10 - < 20					≥ 10 - < 20	
			< 10					< 10	A3				< 10	
		< 3	≥ 20				< 3	≥ 20				< 3	≥ 20	
			≥ 10 - < 20					≥ 10 - < 20					≥ 10 - < 20	
			< 10					< 10					< 10	
	< 5	≥ 5	≥ 20			< 5	≥ 5	≥ 20			< 5	≥ 5	≥ 20	
			≥ 10 - < 20					≥ 10 - < 20					≥ 10 - < 20	
			< 10					< 10					< 10	
		≥ 3 - < 5	≥ 20				≥ 3 - < 5	≥ 20				≥ 3 - < 5	≥ 20	
			≥ 10 - < 20					≥ 10 - < 20					≥ 10 - < 20	
			< 10					< 10					< 10	
		< 3	≥ 20				< 3	≥ 20				< 3	≥ 20	
			≥ 10 - < 20					≥ 10 - < 20					≥ 10 - < 20	
			< 10					< 10					< 10	

Table A 4. Socioeconomic criteria decision tree

APPENDIX B: SENSITIVITY ANALYSIS SUMMARY TABLE FOR CRITERION SOCIOECONOMIC

						Cells in evaluation map			Difference between evaluation map and base map						Changes in evaluation map			
	Weight values					A1	A2	A3	A1		A2		A3		A1 to A2	A2 to A1	A2 to A3	A3 to A2
Change %	SC	TP	UD	CC	UA	# Cells	# Cells	# Cells	# Cells	%	# Cells	%	# Cells	%	# Cells	# Cells	# Cells	# Cells
-20	0.3486	0.3090	0.1657	0.1199	0.0568	5127129	9672375	1486831	112329	2.24	507487	5.54	-619816	-29.42	120635	232964	4803	624619
-19	0.3529	0.3070	0.1646	0.1191	0.0564	5127129	9677178	1482028	112329	2.24	512290	5.59	-624619	-29.65	120635	232964	0	624619
-18	0.3573	0.3049	0.1634	0.1183	0.0560	5127243	9630765	1528327	112443	2.24	465877	5.08	-578320	-27.45	120521	232964	0	578320
-17	0.3616	0.3028	0.1623	0.1175	0.0557	5127243	9630765	1528327	112443	2.24	465877	5.08	-578320	-27.45	120521	232964	0	578320
-16	0.3660	0.3008	0.1612	0.1167	0.0553	5127243	9630732	1528360	112443	2.24	465844	5.08	-578287	-27.45	120521	232964	0	578287
-15	0.3703	0.2987	0.1601	0.1159	0.0549	5134306	9623669	1528360	119506	2.38	458781	5.01	-578287	-27.45	113441	232947	0	578287
-14	0.3747	0.2966	0.1590	0.1151	0.0545	5134306	9623669	1528360	119506	2.38	458781	5.01	-578287	-27.45	113441	232947	0	578287
-13	0.3791	0.2946	0.1579	0.1143	0.0541	5140604	9570133	1575598	125804	2.51	405245	4.42	-531049	-25.21	107143	232947	0	531049
-12	0.3834	0.2925	0.1568	0.1135	0.0538	5140604	9221077	1924654	125804	2.51	56189	0.61	-181993	-8.64	107143	232947	0	181993
-11	0.3878	0.2904	0.1557	0.1127	0.0534	5140604	9221077	1924654	125804	2.51	56189	0.61	-181993	-8.64	107143	232947	0	181993
-10	0.3921	0.2884	0.1546	0.1119	0.0530	5228841	9132840	1924654	214041	4.27	-32048	-0.35	-181993	-8.64	18906	232947	0	181993
-9	0.3965	0.2863	0.1535	0.1111	0.0526	4995894	9365786	1924655	-18906	-0.38	200898	2.19	-181992	-8.64	18906	0	0	181992
-8	0.4008	0.2842	0.1524	0.1103	0.0522	5011085	9350595	1924655	-3715	-0.07	185707	2.03	-181992	-8.64	3715	0	0	181992
-7	0.4052	0.2822	0.1513	0.1095	0.0519	5011085	9350595	1924655	-3715	-0.07	185707	2.03	-181992	-8.64	3715	0	0	181992
-6	0.4096	0.2801	0.1501	0.1087	0.0515	5011085	9342141	1933109	-3715	-0.07	177253	1.93	-173538	-8.24	3715	0	0	173538
-5	0.4139	0.2780	0.1490	0.1079	0.0511	5011085	9342141	1933109	-3715	-0.07	177253	1.93	-173538	-8.24	3715	0	0	173538
-4	0.4183	0.2760	0.1479	0.1071	0.0507	5011085	9342141	1933109	-3715	-0.07	177253	1.93	-173538	-8.24	3715	0	0	173538
-3	0.4226	0.2739	0.1468	0.1063	0.0503	5011602	9341624	1933109	-3198	-0.06	176736	1.93	-173538	-8.24	3198	0	0	173538
-2	0.4270	0.2718	0.1457	0.1055	0.0500	5011602	9168086	2106647	-3198	-0.06	3198	0.03	0	0.00	3198	0	0	0
-1	0.4313	0.2698	0.1446	0.1047	0.0496	5014800	9164888	2106647	0	0.00	0	0.00	0	0.00	0	0	0	0
0	**0.4357**	**0.2677**	**0.1435**	**0.1039**	**0.0492**	**5014800**	**9164888**	**2106647**	**0**	**0.00**	**0**	**0.00**	**0**	**0.00**	**0**	**0**	**0**	**0**
1	0.4401	0.2656	0.1424	0.1031	0.0488	5076442	9103246	2106647	61642	1.23	-61642	-0.67	0	0.00	0	61642	0	0
2	0.4444	0.2636	0.1413	0.1023	0.0484	5076442	9103246	2106647	61642	1.23	-61642	-0.67	0	0.00	0	61642	0	0
3	0.4488	0.2615	0.1402	0.1015	0.0481	5079695	9099993	2106647	64895	1.29	-64895	-0.71	0	0.00	0	64895	0	0
4	0.4531	0.2594	0.1391	0.1007	0.0477	5105304	8732381	2448650	90504	1.80	-432507	-4.72	342003	16.23	0	90504	347764	5761
5	0.4575	0.2574	0.1380	0.0999	0.0473	5105304	8732381	2448650	90504	1.80	-432507	-4.72	342003	16.23	0	90504	347764	5761
6	0.4618	0.2553	0.1369	0.0991	0.0469	5105304	8732381	2448650	90504	1.80	-432507	-4.72	342003	16.23	0	90504	347764	5761
7	0.4662	0.2532	0.1357	0.0983	0.0465	3696875	9606469	2982991	-1317925	-26.28	441581	4.82	876344	41.60	1408429	90504	882105	5761
8	0.4706	0.2512	0.1346	0.0975	0.0462	3707247	9269680	3309408	-1307553	-26.07	104792	1.14	1202761	57.09	1408429	100876	1208522	5761
9	0.4749	0.2491	0.1335	0.0967	0.0458	3707247	9269680	3309408	-1307553	-26.07	104792	1.14	1202761	57.09	1408429	100876	1208522	5761
10	0.4793	0.2470	0.1324	0.0959	0.0454	3707247	9269678	3309410	-1307553	-26.07	104790	1.14	1202763	57.09	1408429	100876	1208524	5761
11	0.4836	0.2450	0.1313	0.0951	0.0450	3807169	9169756	3309410	-1207631	-24.08	4868	0.05	1202763	57.09	1408429	200798	1208524	5761
12	0.4880	0.2429	0.1302	0.0943	0.0446	3866002	9110923	3309410	-1148798	-22.91	-53965	-0.59	1202763	57.09	1408429	259631	1208524	5761
13	0.4923	0.2408	0.1291	0.0935	0.0443	3866002	9110899	3309434	-1148798	-22.91	-53989	-0.59	1202787	57.09	1408429	259631	1208548	5761
14	0.4967	0.2388	0.1280	0.0927	0.0439	3866901	9110000	3309434	-1147899	-22.89	-54888	-0.60	1202787	57.09	1408429	260530	1208548	5761
15	0.5011	0.2367	0.1269	0.0919	0.0435	3866901	9110000	3309434	-1147899	-22.89	-54888	-0.60	1202787	57.09	1408429	260530	1208548	5761
16	0.5054	0.2346	0.1258	0.0911	0.0431	3877706	8550181	3858448	-1137094	-22.67	-614707	-6.71	1751801	83.16	1408429	271335	1757562	5761
17	0.5098	0.2326	0.1247	0.0903	0.0427	3865578	8562309	3858448	-1149222	-22.92	-602579	-6.57	1751801	83.16	1434727	285505	1757562	5761
18	0.5141	0.2305	0.1236	0.0895	0.0424	3865578	8562309	3858448	-1149222	-22.92	-602579	-6.57	1751801	83.16	1434727	285505	1757562	5761
19	0.5185	0.2284	0.1224	0.0887	0.0420	3865578	8562309	3858448	-1149222	-22.92	-602579	-6.57	1751801	83.16	1434727	285505	1757562	5761
20	0.5228	0.2264	0.1213	0.0879	0.0416	3912383	8515504	3858448	-1102417	-21.98	-649384	-7.09	1751801	83.16	1434727	332310	1757562	5761

Table B 1. Sensitivity analysis summary table generated for criterion Socioeconomic. Topography (TP), Urban Density (UD), Climate Conditions (CC), Urban Accessibility (UA)

APPENDIX C: LAND SUITABILITY MAPS FOR UA BASED ON ML MODELS

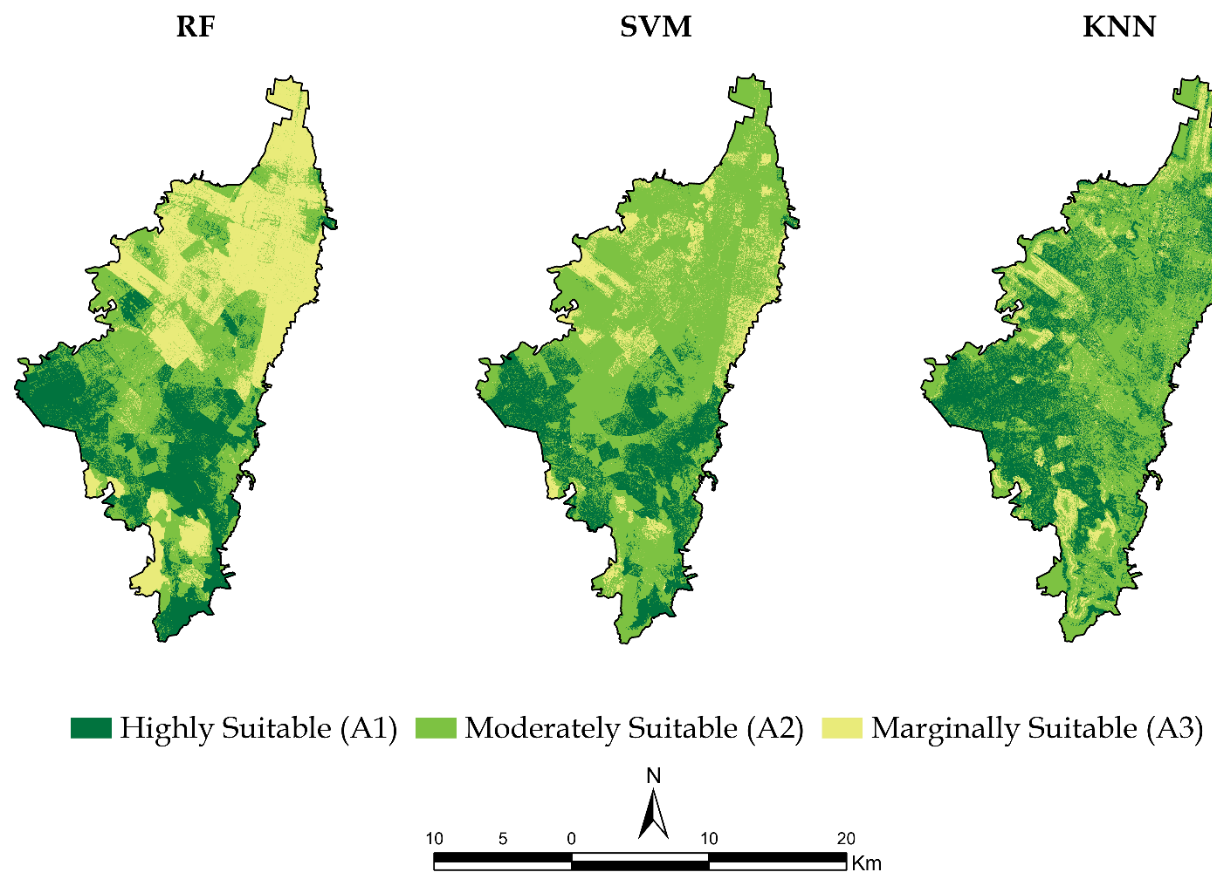

Figure C 1. UA Suitability maps obtained from Random Forest (RF), Support Vector Machine (SVM) and K Nearest Neighbors (KNN) classification models

APPENDIX D: AREAS OF SUITABILITY LAND FOR UA AVAILABLE IN PUBLIC OPEN SPACES

Approach	Public Category	Highly Suitable (A1)			Moderately Suitable (A2)			Marginally Suitable (A3)		
		Cells	Ha	%	Cells	Ha	%	Cells	Ha	%
Expert Knowledge	***Total***	***378701***	***946.8***	***24%***	***993542***	***2484***	***63%***	***208101***	***520.3***	***13%***
	Park	257258	643.1	68%	495334	1238.3	50%	85332	213.3	41%
	Educational Building	71545	178.9	19%	312204	780.5	31%	86186	215.5	41%
	Security Building	12982	32.5	3%	48259	120.6	5%	19145	47.9	9%
	Religious Building	5838	14.6	2%	15820	39.6	2%	3333	8.3	2%
	Medical Building	9796	24.5	3%	24592	61.5	2%	5766	14.4	3%
	Government Building	18364	45.9	5%	69435	173.6	7%	7756	19.4	4%
	Library	2918	7.3	1%	27898	69.7	3%	583	1.5	0.3%
Data Knowledge	***Total***	***283307***	***708.3***	***18%***	***656655***	***1641.6***	***42%***	***640382***	***1601.0***	***41%***
	Park	187019	467.5	66%	335822	839.6	51%	315083	787.7	49%
	Educational Building	63309	158.3	22%	165056	412.6	25%	241570	603.9	38%
	Security Building	4510	11.3	2%	43185	108.0	7%	32691	81.7	5%
	Religious Building	3684	9.2	1%	3844	9.6	1%	17463	43.7	3%
	Medical Building	16759	41.9	6%	16867	42.2	3%	6528	16.3	1%
	Government Building	6146	15.4	2%	62543	156.4	10%	26866	67.2	4%
	Library	1880	4.7	1%	29338	73.3	4%	181	0.5	0.03%
Combine	***Total***	***160421***	***401.1***	***20%***	***464335***	***1160.8***	***58%***	***172425***	***431.1***	***22%***
	Park	117674	294.2	73%	227653	569.1	49%	74619	186.5	43%
	Educational Building	34232	85.6	21%	122407	306.0	26%	70224	175.6	41%
	Security Building	2113	5.3	1%	32827	82.1	7%	17111	42.8	10%
	Religious Building	1568	3.9	1%	2525	6.3	1%	3022	7.6	2%
	Medical Building	1757	4.4	1%	7849	19.6	2%	2743	6.9	2%
	Government Building	1952	4.9	1%	43995	110.0	9%	4665	11.7	3%
	Library	1125	2.8	1%	27079	67.7	6%	41	0.1	0.02%

Table D 1. Areas of suitability land for UA available in public open spaces

APPENDIX E: ESTIMATED AREAS OF URBAN CROPS (TONS) FOR THE PROPOSED SCENARIOS

Approach	Public Category	Lettuce			Chard			Chickpeas			Cabbage			Aloe Vera		
		1st S	2nd S	3rd S	1st S	2nd S	3rd S	1st S	2nd S	3rd S	1st S	2nd S	3rd S	1st S	2nd S	3rd S
Expert Knowledge	***Total***	**686**	**2059**	**3432**	**2575**	**7726**	**12876**	**625**	**1875**	**3124**	**1288**	**3865**	**6442**	**57**	**170**	**284**
	Park	466	1399	2331	1749	5248	8747	424	1273	2122	875	2626	4376	39	116	193
	Educational Building	130	389	648	487	1460	2433	118	354	590	243	730	1217	11	32	54
	Security Building	24	71	118	88	265	441	21	64	107	44	132	221	2	6	10
	Religious Building	11	32	53	40	119	198	10	29	48	20	60	99	1	3	4
	Medical Building	18	53	89	67	200	333	16	48	81	33	100	167	1	4	7
	Government Building	33	100	166	125	375	624	30	91	152	62	187	312	3	8	14
	Library	5	16	26	20	60	99	5	14	24	10	30	50	0	1	2
Data Knowledge	***Total***	**513**	**1540**	**2567**	**1926**	**5779**	**9632**	**467**	**1402**	**2337**	**964**	**2891**	**4819**	**42**	**127**	**212**
	Park	339	1017	1695	1272	3815	6359	309	926	1543	636	1909	3181	28	84	140
	Educational Building	115	344	574	431	1292	2153	104	313	522	215	646	1077	9	28	47
	Security Building	8	25	41	31	92	153	7	22	37	15	46	77	1	2	3
	Religious Building	7	20	33	25	75	125	6	18	30	13	38	63	1	2	3
	Medical Building	30	91	152	114	342	570	28	83	138	57	171	285	3	8	13
	Government Building	11	33	56	42	125	209	10	30	51	21	63	105	1	3	5
	Library	3	10	17	13	38	64	3	9	16	6	19	32	0	1	1
Combine	***Total***	**291**	**872**	**1454**	**1091**	**3273**	**5454**	**265**	**794**	**1323**	**546**	**1637**	**2729**	**24**	**72**	**120**
	Park	213	640	1066	800	2401	4001	194	582	971	400	1201	2002	18	53	88
	Educational Building	62	186	310	233	698	1164	56	169	282	116	349	582	5	15	26
	Security Building	4	11	19	14	43	72	3	10	17	7	22	36	0	1	2
	Religious Building	3	9	14	11	32	53	3	8	13	5	16	27	0	1	1
	Medical Building	3	10	16	12	36	60	3	9	14	6	18	30	0	1	1
	Government Building	4	11	18	13	40	66	3	10	16	7	20	33	0	1	1
	Library	2	6	10	8	23	38	2	6	9	4	11	19	0	1	1

Table E. Estimation of urban crop production in Tons for the proposed scenarios(S)