Addis Ababa University

College of Natural Sciences

Design of a Model for Service Facility Site Selection: Case of Addis Ababa

Firesew Feyiso Woldesillasie

A Thesis Submitted to the Department of Computer Science in Partial Fulfillment for the Degree of Master of Science in Computer Science

Addis Ababa, Ethiopia

June, 2018

Addis Ababa University

College of Natural Sciences

Firesew Feyiso Woldesillasie

Advisor: Solomon Atnafu (PhD)

This is to certify that the thesis prepared by Firesew Feyiso, titled: *Design of a Model for Service Facility Site Selection: Case of Addis Ababa* and submitted in partial fulfillment of the requirements for the Degree of Master of Science in Computer Science complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Name	Signature	Date

Advisor: ______________________________________

Examiner: ______________________________________

Examiner: ______________________________________

Abstract

An increase in the number of residence with an increase in urbanization results a growing demand of quality services. The geographical relationship of infrastructures,economic and political aspect of urban region is requred to manage and plan the activities of service facilities. Lack of all-inclusive site selection with the consideration of geographical relationship of the existing infrastructure, population distribution and density, natural and artificial barriers, legal and political aspect of the location result improper distribution and allocation of resources for the public and miss use of the existing services facilities.

Site selection in Addis Ababa usually depends on the existence of free and cheap land. The purpose of this study is to design and develop a model that can identify and support decision makers to select optimal service facility site using suitability analysis and data mining outputs for Addis Ababa. Our model employed unsupervised classification of data mining called spatial clustering and suitability analysis using GIS-based multi Criteria Analysis for optimal site selection. First, suitability of the study area will be identified using carefully selected and weighted factor and constraint criteria by pair wise comparison using analytical hierarchical process, which is the systematic engineering method to quantitatively analyze non-quantitative objects. Then spatial data mining using road data identifies clusters and cluster center. Finally suitability map and data mining results are used to select candidate location and optimal site selection using different kind of vector and image maps.

The model implemented using C# and is tested for hospital site selection using Nifasilik Lafto sub city, Addis Ababa and suitability map of Addis Ababa for hospital site selection. The overall optimal site selection accuracy is tested and evaluated by experts and it has been found that different candidate site and optimal site is selected. The overall performance of the proposed model from summary of validation report is 85.2%. And the system can provide adequate justification or reason for the selected site is about 88% effective.

Keywords: Spatial Data Mining , Suitability Analysis, Optimal Site Selection, Candidate Location, GIS-based multi Criteria Analysis, Addis Ababa

Dedication

To my uncle Abebe Mammo

ACKNOWLEGEMENT

First and for most, I would like to glorify the almighty God you have done so much for me. Then, I would like to forward a very much grateful thank to my advisors **Dr. Solomon Atinafu** for his continuous and unreserved support, guidance, constructive comments and suggestion. It is with his continuous follow up, encouragement and advice that this work came to realization. I appreciate his dedication from the beginning to the end to teach me for the first time how to conduct and write a research.

I would also like to take this opportunity to express my profound gratefulness to my families for their continuous support. My sincere thanks also go to my friend **Mr. Sintayehu Zekariyas** , for his advice, guidance and constructive comments from the beginning to the end of the work and **Mr. Birhanu Kifle** who advised me on GIS related issues. I also like to thank urban design department for their cooperation in advice and evaluation of my work. I also want to thank Graduate Committee of Addis Ababa University Computer Science Department for giving me the permission to commence this research work. Since this thesis is the cumulative results of the two years learning, I would like to thank all the staff members of the department of Computer Science involved in the process as well as my classmates for working together in different projects and assignments in harmony.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AHP	- Analytical Hierarchical Processing
AICOE	- Artificial Immune Clustering with Obstacle Entity
CLARA	- Clustering Large Applications
CLARANS	- Clustering Large Applications based on Randomized Search
DEM	- Digital Elevation Model
DMS	- Degree Minutes Second
DRSA	- Dominance Based Rough set Approach
ELECTRE	- Elimination and Choice Expressing the Reality
GC	- Gregorian Calendar
GIS	- Geographic Information System
IT	- Information Technology
MAUT	- Multi-utility Attribute Utility Theory
MCA	- Multi Criteria Analysis
PAM	- Partitioning Around Medoids

PD	- Perimeter Distance
P2D	- Path to Point Distance
PROMETHEE	-Preference Ranking Organization Method for Enrichment of Evaluation
ROM	- Random Order Methods
SDM	- Spatial Data Mining
SPD	- Shortest Path Distance
UTC UTM WGS USGS	-Coordinated Universal Time -Universal Transverse Mercator -World Geodetic System - United States Geological Survey

CHAPTER ONE

INTRODUCTION

1.1 Background

In the last three decades, the rapid population growth of Addis Ababa has been putting tremendous pressures on the city. In 1992, the population was nearly 1.4 million. In 2008 it was more than 4 million [**1**].This high growth has been putting a great pressure on both social care and physical infrastructures. Increasing in urban economies and size needs additional service facility establishment to provide a better service. In Addis Ababa the methods of location selections of service facilities are not optimal according to our discussion with urban design. Usually it is based on the existence of free land within the administrative boundaries of sub cities. Site selection needs a holistic view between the geographical relationship of the existing infrastructure, population distribution and density, natural and artificial barriers, legal and political aspect of the location for planning, distribution and allocation of resources for the public. Urban planners need techniques to select optimal service facility site and location selection for public use because planning is not just a practice of monitoring the space, but rather is a skill or an art with a broad influence on society.

The quality of any services is not only determined by the number, quality of staff and the infrastructure in which the facilities it contains but also the reachability of its location, environmental, economic and legal aspects also have significant influence on the services it provides. The spatial connectivity found in the city should be considered during the site selection process because the road transport network is mainly planned to connect local resources and people in distant markets and population centers [2].

Addis Ababa has a network structure that we can represent in different form and can help us in the service facility site selection. Decision on site selection could also be reviewed and proven based on the analytic hierarchy process (AHP) and statistical analysis evaluation for site selection. The work in [3] stated that the result of industrial wastes in river, soil, air and water pollution is poorly managed and are becoming growing concerns in Addis Ababa.

So, we used site selection based on the analytic hierarchy process (AHP) with road network analysis because AHP establishes decision weights for alternatives by organizing objectives, criteria and sub criteria in a hierarchic structure and transfer qualitative criterion measures into quantitative or convert subjective parameter measures into computable or numerical for service facility location selection

1.2 Motivation

In a city as big as the capital of a country, there is a need to consider the basic steps and procedures in urban planning to provide a standard service, this includes the selection of sites for public services that consider the constraints on the cost of transportation [4] and suitability of the site for the associated facilities. The geographical relation of objects in a city such as buildings, streets, rivers, bridges and land structure can help us to identify where the optimal location is by extracting knowledge from the available data using proper spatial data mining techniques and suitability analysis of the selected location or candidate site. The use of such methods to select an appropriate location could be handled using different kinds of applications. Policy makers and urban planners can use such applications to identify what exists and what will be the future site to develop new guidelines and policies. We observed and discussed site selection methods in Addis Ababa design office, the office identify and approve the suitability of the identified location based on number of population served and availability of land reserved but using both suitability and data mining for service site selection at the same time is not practiced. Since such practices are not used in Addis Ababa, designing a service facility site selection model using different geospatial analysis methods to leverage their advantage is the motivating factor for picking this topic.

Accordingly, this work will have the following significances:

- It helps to facilitate the process of building service facility by selecting an appropriate location within a short period of time.
- It helps to analyze the number of parcels which could be displaced in advance.
- It helps to analyze the distribution of the existing urban service facilities ahead of an investment.

- It helps for decision makers to decide the optimum site selection based on existing spatial data by managing many conflicting criteria.
- It helps urban planners to use standard and procedures for site selection and avoids corruption.
- Policy makers can make use of it to develop new guidelines and policies.

1.3 Statement of the Problem

According to the standard set by the ministry of health, one district hospital is for 250,000 people, one regional hospital for 1,000,000 people and one specialized hospital for 5,000,000 million people as a result, the projected demand of hospitals for Addis Ababa for the year 2020 will be a total of 20 additional hospitals [3].To fulfill the above demand, encouraging result has not only been achieved by the government, the private sectors have also been playing a substantial role in improving the physical coverage of service facilities. But practical access remains a challenge as the private service institutions are not affordable for the majority of the population.

Increasing urbanization with an increase in the number of residence needs services with optimal service facility site selection which results increase in accessibility of the selected site. In the past, site selection was based on economic and technical criteria. But, today cities of the world experiencing different site selection methods, as the authors in [5,6] stated decision making is based on numerous data concerning the problem and site selection needs a higher degree of sophistication to satisfy social and environmental requirements, which are enforced by legislations and government regulations. The existing site selection models are much advanced than manual way of site selection of Addis Ababa and didn't practice before for Addis Ababa. Thus, existing service facility sites of Addis Ababa are just selected and constructed without all-inclusive observation of the existing population distribution with the geographical barriers and man-made constraints [7]. As a result, residents of the city, travel long distances, waits a long time for emergency services and waste of time because of more crowds for one specific service area. Government and private service providers are also facing the challenges of delivering emergency and other services on time. Our model bridges the gap between Addis Ababa and some other developed cities for service facility site selection because, it computes the centrality of the selected site using networked distance rather than Euclidean distance and considers all obstacles and facilitators exist in urban area in addition, it use the possible advantage of suitability analysis like some other models in the world.

To this end, this work attempts to explore and answer the following research questions:

- Which data mining techniques and which suitability analysis methods are appropriate to select the optimum service facility site?
- Where in the city is the optimum site for the next service facility?

1.4 Objective

General Objective

The general objective of this work is to design and develop a model that can identify and support decision makers to select optimal service facility site using suitability analysis and data mining outputs in case of Addis Ababa.

Specific Objective

In order to attain this general objective, the following specific tasks will be performed.

- Review of literatures and related work
- Collection of relevant data and documents of service facility site selection
- Study and analyze the nature of service facility site selection
- Identify and represent the existing geographical barriers and facilitators.
- Prepare data for analysis and spatial clustering by cleaning, extracting, and transforming it into a format suitable for the spatial data mining and suitability analysis
- Select the necessary algorithms or tools for spatial clustering analysis that help in implementing the designed model.
- Select suitability analysis method used for site selection.
- Clustering the data that helps to identify the appropriate location.
- Design a model that can be used for service facility site selection.
- Develop a prototype system based on the designed model.
- Test and validate the designed system

1.5 Methodology

To achieve the general objectives of the research, the following methodologies will be used.

1.5.1 Literature Review

Different literatures which are considered to be relevant for the research will be reviewed and some of the concepts and models will be adopted for our work. Since our research work is on optimal service facility site selection it contains different types of domain areas like GIS based suitability analysis and spatial data mining: the concepts of site selection, knowledge discovery methods using the existing spatial data and techniques will review in order to understand and extract novel, understandable and usable knowledge from the existing data. Urban network analysis, spatial clustering and road network analysis will review in order to identify the existing urban objects, spatial relationship, characteristics and representation of urban road used for spatial data mining location identification and parameters used for the analysis. We will review population condition in Addis Ababa, identify existing service facilities and parameters used for service facility site selection procedures and standards to understand the legal and political issues which help us to integrate the site selection criterion and rules during suitability analysis. We will also review relevant work on GIS and GIS based site selection to understand the characteristic, type and the way to manage spatial data used.

1.5.2 Data Collection

The inputs for the design of a model for service facility site selection decision making are both spatial data mining and suitability analysis outputs. To achieve our general and some other specific objectives, we use vector data from the open street map, Addis Ababa satellite image data will be downloaded from Google earth using Addis Ababa extent. The vector data which we can't find from the open street will be digitized from the downloaded image which means, for university area boundary and disposal area ('koshe') boundary, we extracted from the satellite image using ArcGIS editor tool.

1.5.3 Analysis and Design

To design the model of service facility site selection, different GIS based site selection system models and suitability analysis models developed before will be studied. To successfully accomplish our system implementation, C# programming language and GIS software applications will be in use.

1.5.4 Evaluation

By comparing the accessibility and centrality of the study area using our model we can evaluate the performance and accuracy of the developed model. After mining the needed spatial knowledge and overlapping with suitability analysis result, decision makers and domain experts will evaluate the accuracy and performance of result because the final validation for the model is based on the spatial knowledge and suitability output. The relative validations for both outputs handle by using the tools used for domain experts. Finally, the conclusions and recommendations will be driven from the evaluation results.

1.6 Scope and Limitations

The real population distribution of any country is based on the quality and methods of census data collection. If the collected data contains spatial information for each household, it is easy to represent the real spatial distribution of the residents. We will replace the overall spatial relationship of the Addis Ababa residence by using the existing road junctions (vertex) and road edges in order to, identify population based on distance and neighborhood relationship rather than their distribution because the existing collected census data is paper based and is not geo-referenced. Site selection of the area is therefore based on road junction and edges that represent the spatial relationship of residence and the possible geographic barriers will be considered using road (street + bridge) network dataset as spatial constraint.

1.7 Application of Results

The proposed system in this work will use spatial relationship and distribution of infrastructures for human activities and movements to the select service facility site. After the discovery of knowledge using group analysis method and suitability analysis, results will be overlap and further analyze by decision makers. The implementation of our model can measure the surrounding area of cluster center point and helps to identify sites suitable for service facility close to the center point of the beneficiaries (clusters), it will also show parcels affected by the selected site and adjusting the selected sites around the selected area will also easy. User's requirement could depend on types of facility they select. Before investing in the selected area, private investors or government agencies can analyze the pro and cons of the selected area based on their criteria. Policy makers can develop new guideline and policies by analyzing the current existing service facility distribution moreover, city planners will conduct their activity based on the guideline and government policies using our application and the future need of the government.

1.8 Organization of the Thesis

The rest of the thesis is organized as follows: Chapter Two presents the literature review of site selection, GIS, GIS based site selection, spatial data mining and urban network analysis. Chapter Three discusses the related works in spatial data mining for facility location selection and GIS based multi criteria analysis for specific facility site selection and hospital site selection analysis. Chapter Four presents the model for the proposed site selection and discusses the function of each component and their relation. The implementation detail of the model such as tools, techniques and methods used to develop the system will describe in Chapter Five. Chapter Five presents the evaluation and results analysis obtained from the proposed system along with their interpretations and the reasons behind each of the results. Finally, in Chapter Six the conclusions drawn the contributions of the work, recommendations, and future research directions are given.

CHAPTER TWO

LITERATURE REVIEW

In this chapter, main research works related to the subject are reviewed. The concept of site selection, GIS & GIS based service facility site selection and spatial data mining**,** spatial clustering; urban network analysis has been reviewed. The geographic condition of Addis Ababa city, land use type, the condition of population distribution in the Addis Ababa city and the service facilities and their parameters for each service facility site selection have been also considered because it is the basis for the methods and techniques for the design of a model for service facility site selection.

2.1 Description of the Case Study Area

We used Addis Ababa the capital city of Ethiopia for our study. Addis Ababa, the federal capital of Ethiopia, was founded in 1886. Lies 9°1′48″N Latitude and 38°44′24″E Longitude. The city is located at the mid of the country and surrounded by Oromiya region, at an altitude ranging from 2,100 meters at 'Akaki' in the south up to 3,000 (9,800 ft) meters at 'Entoto' Hill in the North. Its time zone is categorized in East Africa Time (UTC+3).

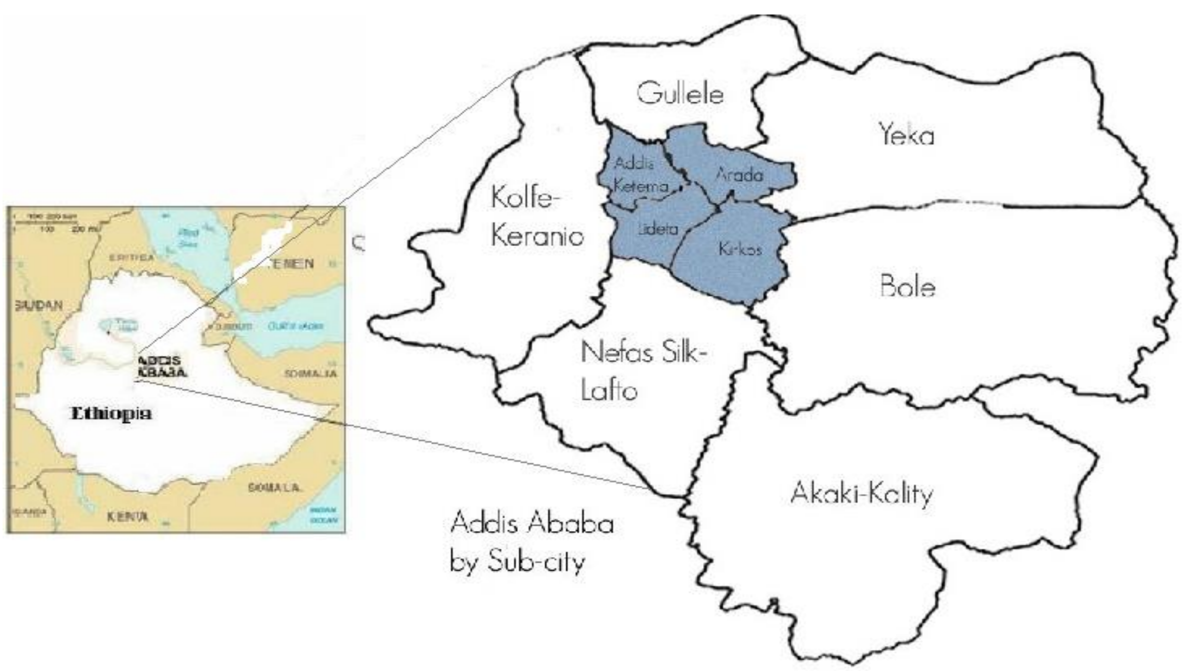

Figure 2.1 : Addis Ababa (source: www.researchgate.net/figure/Map-of-Addis-Ababa-City - fig1_281460707)

The city occupies a total area of 540 Sq.km [**1**].As shown in Figure 2.2 Addis Ababa contains ten sub cities.

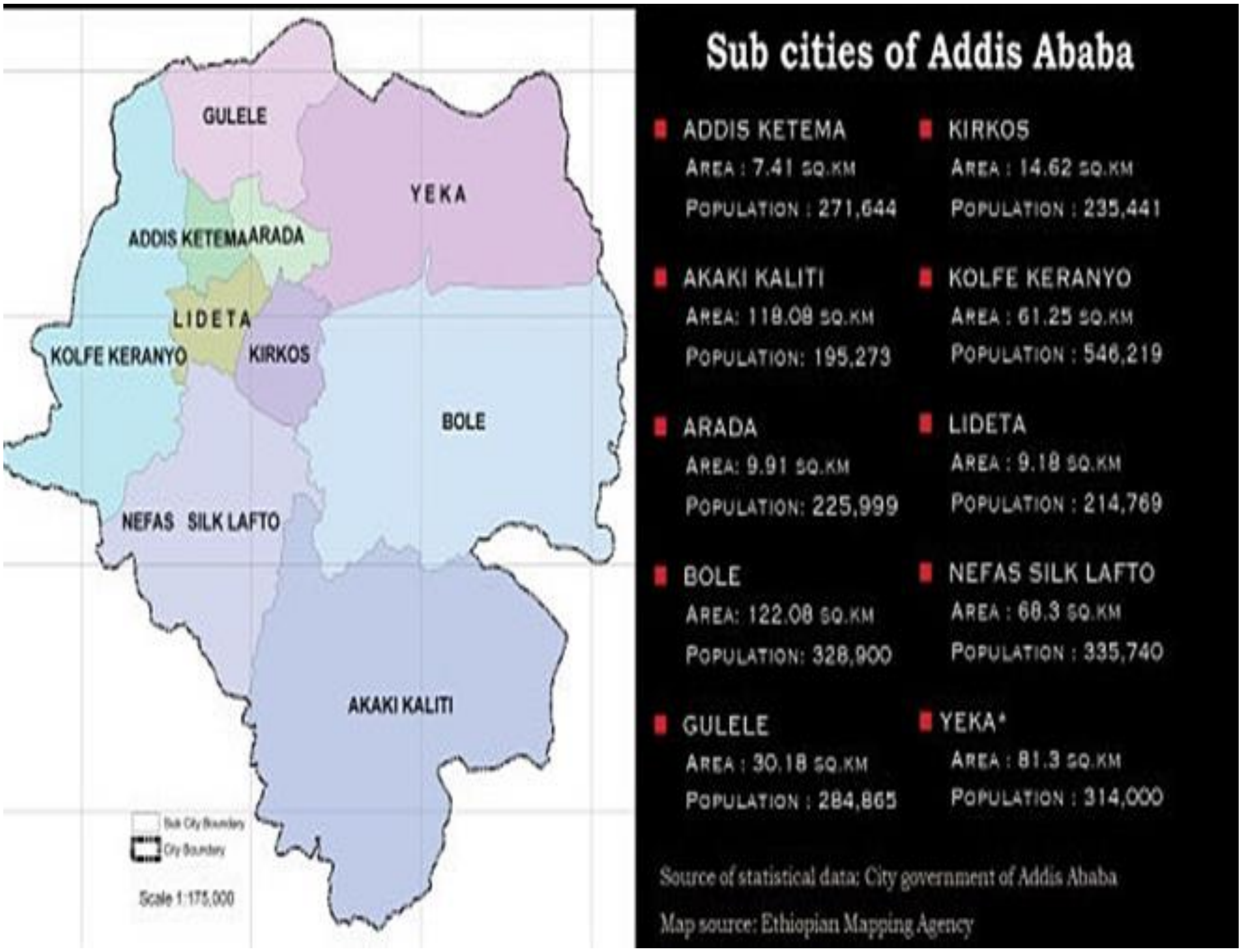

Figure 2.2 : Sub Cities of Addis Ababa [6]

2.2 Land use Type of Addis Ababa

Addis Ababa has different type of land use like other cities. Mainly, it is occupied by residential area, field crop, open space and road as shown in Figure 2.3 [8]

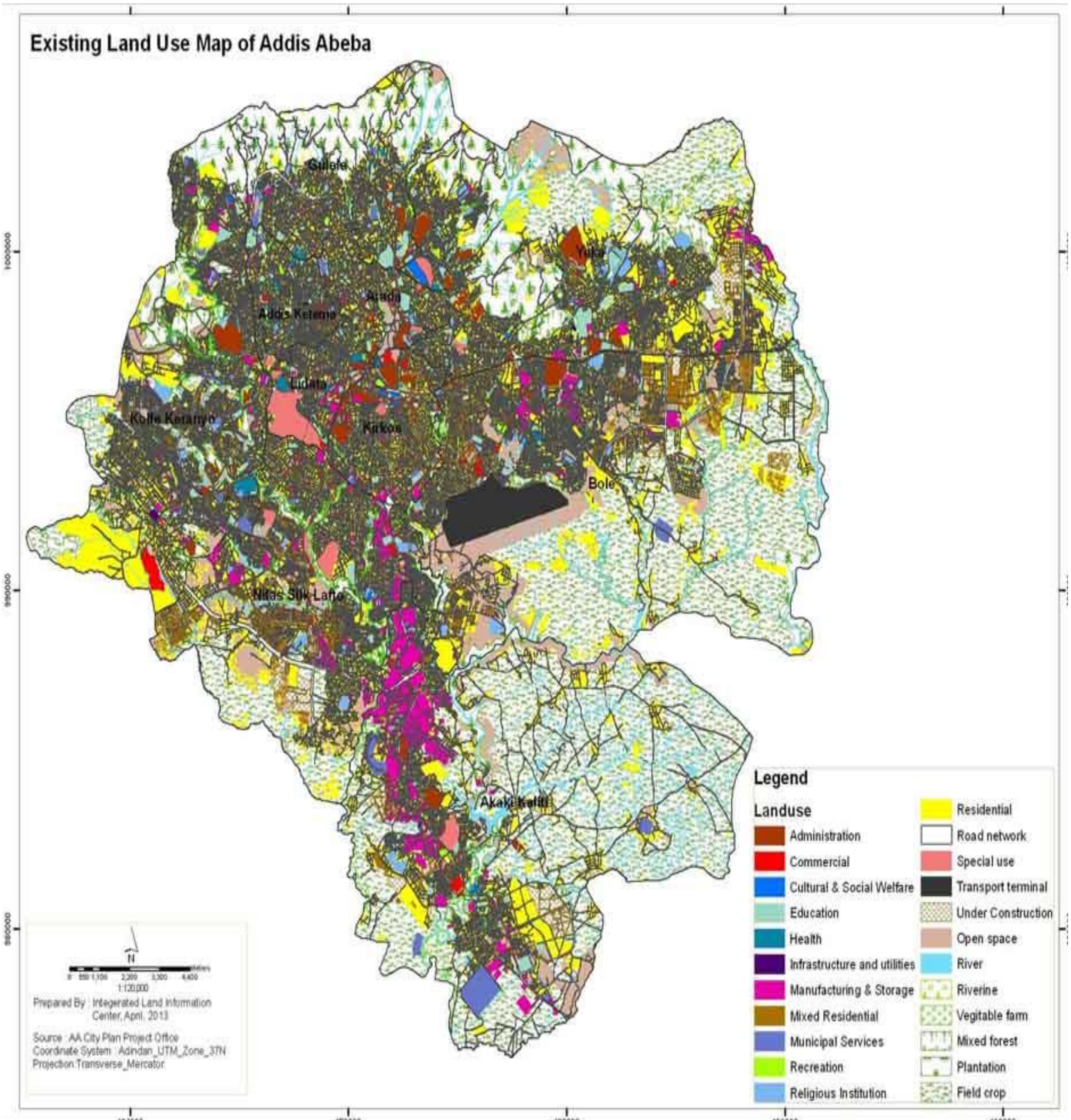

Figure 2.3 : Land Use Type of Addis Ababa. **(**Source: www.ilic.gov.et/index.php/en/land-use)

2.3 Conditions of Population

Addis Ababa, the capital of Ethiopia and the diplomatic center of Africa, is one of the fastest growing cities in the continent and hosts 30 percent of the urban population of Ethiopia. Its population has nearly doubled every decade. In 1984 the population was 1, 412, 575, in 1994 it was 2,112, 737, and it is currently thought to be 4 million. UNHABITAT estimates that this number will continue to rise, reaching 12 million in 2024 [**3**,**1**] .The population distribution and density of each sub cities depend on the housing condition, parcel ownership type and city land administration land use plan. We used 2007GC census data which is not latest. The population of Addis Ababa in 2018 is estimated to be 6.6 Million + 0.578 Million = 7.178 Million and the estimated values differ from different sources. Census has been extended because of the crisis situation in the country so 'Lideta' sub city,' Yeka' sub city & 'Arada' sub city are relatively more dense than the other sub cities as shown in Table 2.1 [8].

Table 2.1**:** Population Density of Addis Ababa Sub cities

No	Sub City Name	Area in km2	Population Size	Population Density (peoples/Sq.Km
1	Yeka	82.13	346484	4218.67
2	Nifas Silk	58.76	316108	5379.344
3	Lideta	9.18	201613	21952.47
4	Kolfe Keranyo	63.48	415647	6752.381
5	Kirkos	14.65	220991	15087.18
6	Gulele	31.19	267381	8572.377
7	Bole	118.5	308714	2605.291
8	Arada	9.5	212009	22316.7
9	Akaki kaliti	123.46	181202	1467.686
10	Addis Ketema	8.64	268099	31029.9
Total		519.49	2,738,248	5,271

2.4 Site Selection

Site selection is the practice and process of identifying the new facility location and specific site for business or government use. The selection of a site indicated in [9] includes both location and site selection. Location denotes to a general area within a city, while the site is a specific piece of parcels. In the past, site selection was based almost purely on economic and technical criteria. Today, a higher degree of sophistication is expected as pointed out in [5], government legislations and regulations enforce site selection criteria to satisfy a number of social and environmental requirements because the primary consideration in site selection is an appropriate location; with the objective of "the right care, in the right place, at the right time" [9].

In another word, the process of site selection typically involves two main phases: screening i.e. the identification of a limited number of candidate sites from a broad geographical area based on a range of selection factors and in-depth examination of alternatives to determine the most suitable site [9]. In terms of hospital site selection, many factors should be taken into account such as the population, pollution, existing service centers, geological factors, economic factors, government policies and administrative laws and regulations [10]. Therefore, making good preparations and analysis of the parameters of the site selection is absolutely necessary [11]. The work in [12,11,13] emphasized that an appropriate criteria of site selection help for the optimization of resource allocation, the urban and rural service development and coordination, facilitate service care accessibility, rescue time reduction satisfies and promote the life quality. By providing service

The work in [12] stated service facility site selection from different aspects:

- From the aspect of the government, appropriate service site selection will help to optimize the allocation of resources, matching the provision of service care with the social and economic demands, coordinating the urban and rural service development, and easing social contradiction.
- From aspect of the citizen, proper service site selection will improve access to the service, reduce the time of rescue, satisfy service needs as well as enhance the quality of life.

- From the aspect of the investors and operators of the service, optimum service site selection will definitely be cost saving on capital strategy. Recently, site selection has been deemed necessary for the sustainable development of land use, to solve the issue of competitive demands of space, to avoid undesirable environmental loads, and to ensure profitability of space. Site selection is an important planning process which affects different regions in the economy, the ecology, and the environmental service sectors

2.5 GIS Based Site Selection

Many IT related literatures like the work in [6,14,5] defines GIS as an innovative and integrated technology based on many disciplines such as Computer science, Geography, Cartography, Statistics, Remote sensing, Land surveying and Navigation. GIS consists of five key parts: hardware, software, data, methods and people. The data and information in GIS are geographically referenced (Geo-coded). Referring to the GIS methods, the work in [10,5] concluded that GIS has the basic capabilities of data input, storage, retrieval, output function and more advanced capabilities of spatial data analysis and scientific modeling. The work in [6] shows GIS approaches used in the site selection usually are network analysis, spatial analysis, proximity analysis, Multi Criteria Analysis (MCA) with Analytical Hierarchy Process (AHP) and Rank order method (ROM).

Decision making is based on several and different types of data regarding the problem. As stated in [15,5] managers and some other decision makers' uses an estimate of 80% spatial data or data which have geographical in nature for their decision. Likewise, most decision makers need neighborhood of a spatial object to know the spatial relation and dependencies of existing facilities so that, GIS has the capability to model such dependence easily. The use of GIS defined in [16] allows for viewing and analyzing the effects of locating a proposed facility in the neighborhood of existing facilities/utilities. GIS also has a specific potential role in evaluating the geographic distribution of different services, particularly evaluating the effectiveness of service facility coverage relevant to population. Various GIS analytical tools have been used widely to recognize spatial patterns of distribution of existing service facility and find out the new optimal location of facilities [14].

A GIS based support tools such as urban network analysis, network analysis and spatial statistics integrate environmental considerations into local spatial planning that make GIS tools as a rational solution to produce natural resources and the conservation of areas through the evaluation of environmental sustainability [16]. GIS is useful for manipulating spatial aspects of selecting suitable sites due to the ability to bring together many diverse and complex factors to facilitate development and administrative decisions [12].

2.6 GIS Based Service Facility Site Selection

Choosing service facility location is the main problem for city planners. The problem is wider in developing countries because of fast growing urbanization and poor planning [12]. The work in [10,6,5] clearly stated that, service facility site selection should participate service facility management personnel, engineers, and government officials, social and environmental responsible bodies even though, it results different conflicting ideas and interests. GIS-Based Multi Criteria Analysis (MCA) method is used to transfer qualitative criterion measures into quantitative or convert subjective parameter measures into computable or numerical and make the result more convincing [6].

According to the work in [6] , GIS-based MCA includes two essential parts: factor criteria and constraint criteria. Each of the criteria appears as a map layer, for both factor or constraint criterion. Factor maps are represented as spatial distributions to display the opportunity criteria and the quality of achieving an objective. Constraint maps are limitations or restrictions which prohibit certain elements to be taken into account in the analysis. GIS-based MCA includes two major methods: weighted summations procedures and Boolean overlay operations. Weighted summation means the linear combination or the summation of factor criteria using the given weight of each factor criteria. The final output will be ranked based on the summation values of each cell ranging from 0-255. Boolean overlay means the multiplication of constraint criteria based on the value given to each constrains ‘1’ and ‘0. Value ‘0’ means the area is restricted for facility site selection ‘1’ means it is possible to consider the area.

2.7 Multi-Criteria Analysis (MCA)

Sustainability assessments require the management of a wide variety of information types, constraints and suspicions. Multi criteria analysis (MCA) has been regarded as a suitable set of methods to perform sustainability evaluations as a result of its flexibility and the possibility of facilitating the discussion between stakeholders, analysts and scientist

Different literatures define or represent MCA in different ways with the same general concept. The work in [17] defined MCA as it covers a range of appraisal techniques that have the potential to capture a wide range of impacts that may not be readily valued in monetary terms, especially those relating to social issues. MCA aims to establish references between options by reference to an explicit set of specified objectives and associated criteria for assessing the extent to which these objectives have been achieved. Two of the key advantages of MCA are that it can allow greater stakeholder involvement and provide greater transparency to the decisions being made at all levels of assessment.

MCA covers a range of techniques for assessing decision problems characterized by a large number of diverse attributes where these do not need to be expressed in money terms. At a simple level, there is a range of methods that have been designed to screen out ‘worse’ options and possibly to identify the ‘best’ option, without aggregating information across different attributes. In contrast, some of the more sophisticated techniques are aimed at providing a means for aggregating information into a single indicator of relative performance.

The work in [18] Defines MCA as is a general framework for supporting complex decision-making situations with multiple and often conflicting objectives that stakeholders groups and/or decision-makers value differently. A typical example of a decision-making situation assisted by MCA methods is determination of an appropriate water regulation policy, which has a variety of economic, ecological and social consequences regarded as desirable by some stakeholders .

As stated in [19], at the present moment (2018), about 100 different methods of MCA are in circulation among them, the following are the most frequently used and implemented

- **Multi attribute utility theory (MAUT**)

Multi attribute utility theory is a performance aggregation based approach, which requires the identification of utility functions and weights for each attribute that can then be assembled in a unique synthesizing criterion with the additive and multiplicative aggregations being the most widely applied . The selection of the relevant aggregation procedure requires the verification of various assumptions, and a discussion about this important procedural step can be found in

- **Analytical hierarchy process (AHP)**

The analytic hierarchy process is another approach of the first performance aggregation based approaches and it was introduced with the aim of evaluating tangible and intangible criteria in relative terms by means of an absolute scale

- **Elimination and choice expressing the reality (ELECTRE)**

ELECTRE are preference aggregation based methods, working on pair-wise comparisons of the alternatives. These methods were introduced by Bernard Roy preferences are structured on four elementary binary relations, indifference, preference, weak preference and incomparability

- **Preference ranking organization method for enrichment of evaluations (PROMETHEE)**

PROMETHEE methods are also part of the outranking MCA family and were developed during the early eighties, and are based on a set of prerequisites

the extent of difference between the performance of two alternatives must be accounted for;

1. the scales of the criteria are irrelevant as comparisons are performed on a pairwise base;
2. three cases are possible: alternative a is preferred to alternative b ;alternative a and alternative b are indifferent; alternative a and alternative b are incomparable;
3. the methods should be easily understandable by the decision makers;
4. Weight must be assigned in a flexible manner.

- **Dominance-based rough set approach (DRSA)**

The dominance-based rough set approach is a relatively new technique which can handle classification, choice and ranking problems

Considering the premises, AHP resulted as the most suitable for our work because:

- The application of the AHP allows to decompose the decisional problem
- Structuring hierarchic levels that allow dealing with the decision and reducing its complexity.
- The AHP allows for choosing the most suitable alternative based on quantitative and qualitative evaluations

- The process can be improved and supported by consistency analysis on the evaluations expressed by the decision-maker or the group of decision-makers and the sensitivity analysis allows to check the influence of each criteria on final decision.

2.7 Analytical Hierarchy Process

The work in [20,21,9] defined AHP is:

- The quantitative method for selecting among alternatives based on their relative performance with respect to one or more criteria of interest.
- A simple systematic engineering method to quantitatively analyze non-quantitative objects

The hierarchy of AHP is constructed through pairwise comparisons of individual judgments, rather than attempting to prioritize the entire list of decisions and criteria simultaneously.

The AHP procedure generally involves six steps [22]:

- **Define the unstructured problem, identification of input/output parameters**

In this step the unstructured problem and their characters should be recognized and the objectives and outcomes stated clearly.

- **Representation of a structure by hierarchy**

The first step in the AHP procedure is to decompose the decision problem into a hierarchy that consists of the most important elements of the decision problem. In this step the complex problem is decomposed into a hierarchical structure with decision elements (objective, attributes i.e. criterion map layer and alternatives)

- **Paired comparison between elements at each level**

For each element of the hierarchy structure all the associated elements in low hierarchy are compared in pairwise comparison matrices

- **Calculations of the weight at each level**

Some methods like eigenvalue method are used to calculate the relative weights of elements in each pairwise comparison matrix

- **Test the consistency of each matrix**

In this step the consistency property of matrices is checked to ensure that the judgments of decision makers are consistent

- **Priority of an alternative by a composition of weights**

In last step the relative weights of decision elements are aggregated to obtain an overall rating for the alternatives

Decision is a choice between alternatives. Criterion are some basis for a decision that can be measured and evaluated. It is the evidence upon which a decision is based. Criteria can be of two kinds: factors and constraints.

- A factor is a criterion that enhances or detracts from the suitability of a specific alternative for the activity under consideration. It is therefore measured on a continuous scale.
- A constraint serves to limit the alternatives under consideration. In many cases constraints will be expressed in the form of a Boolean (logical) map: areas excluded from consideration being coded with a '0' and those open for consideration being coded with '1' [6,5]

2.8 Spatial Data Mining

A major difference between data mining in regular relational databases and in spatial databases is that attributes of the neighbors of some object of interests may have an influence on the object and therefore have to be considered as well. The explicit location and extension of spatial objects define implicit relations of spatial neighborhood (such as topology, distance and direction relation) which are used by spatial data mining algorithms [23]. Spatial data mining and knowledge discovery defined in [24,25,26].

- As the efficient extraction of hidden, implicit, interesting, previously unknown, potentially useful, ultimately understandable, spatial or non-spatial knowledge (rules, regularities, patterns, constraints) from incomplete, noisy, fuzzy, random and practical data in large spatial databases.
- It is a confluence of database technology, artificial intelligence, machine learning, probabilistic statistics, visualization, information science, pattern recognition and other disciplines.

Main tasks in SDM are spatial outlier detection, co-location pattern discovery, spatial classification and regression modeling, spatial clustering, and spatial hotspot analysis [26].

2.9 Spatial Clustering

Spatial clustering is a method of grouping a set of spatial objects into clusters so that, objects which grouped in one cluster have high similarity with one another compared to objects in another cluster [10,27,26]. Focus on the spatial dimensions of objects, many clustering algorithms have been developed to undertake the task of spatial clustering even if most developed algorithms have ignored the fact that obstacles such as rivers, lakes, and highways exist in a real world and could thus affect the result or the quality of clustering [10]. In spatial cluster analysis, finding the structure directly from a big dataset without relying on the prior knowledge of spatial hierarchies is the beauty and capability of clustering algorithms in many geographical or spatially related knowledge discovery tasks [24].These spatial clustering methods can be categorized into four broad categories: hierarchical, partitioned, density-based, and grid-based [16]. Clustering algorithms can be generally categorized into Partitioning methods, hierarchical methods; density-based methods and grid-based methods [32].There are many motivations behind selecting partition-based methods for this problem.

First, the hierarchical methods suffer from poor scalability with increasing the number of points. Second, the main advantage of density based methods is their ability to find elongated and non-convex clusters. This is a valuable capability in spatial data mining applications. Nonetheless, this is not useful in the problem of finding best locations for facilities since here the objective is to minimize the customers covering cost. Third, grid-based approaches also suffer from some shortcomings as a possible solution to our problem. The performance of these algorithms relies on many parameters such as the granularity of the lowest level of the grid structure and assumptions on data distribution and also the resulting clusters are bounded horizontally or vertically, but never diagonally. Considering the concept of overall distance must be minimized, the most appropriate algorithms among the above mentioned categories are partitioning methods, in addition, embedding the objective of optimizing service facility establishment; partitioning methods are much less complicated compared to the other.

The above explanations are the main motivations for using partitioned-based approaches in most of the spatial clustering methods. So, we have used partitioning method for optimal service facility site problem.

2.10 Partition Based Clustering

Partition clustering methods determine a partition for dividing a group of points into different clusters, such that the points in a cluster are more similar to each other than two points in different clusters. The technique used by the partitioning clustering algorithm as stated in [26] these methods start with some arbitrary initial clusters and iteratively reallocate points into clusters until a stopping criterion is met. These methods tend to find clusters of spherical shape. K-Means and K-Medoids are commonly used partitioned algorithms. Squared error is the most frequently used criterion functions in partitioned clustering. The recent algorithms in this category include partitioning around medoids (PAM), clustering large applications (CLARA), clustering large applications based on randomized search (CLARANS), and expectation-maximization (EM) [**1**,**26**]

2.11 Urban Network Analysis

Nowadays, cities and towns contain different types of network like transportation network, power grids and water utility networks which currently becoming complex [28] .The road network is one of largely planned transportation networks to connect local properties and people to distant places and community center. Thus, it is important for maintaining and civilizing the quality of life for urban community because it provides support to urban system development [2]. The work in [15] defined Road network as undirected or directed weighted spatial graph as shown in Equation (1) below:

$$G = (V; E; W) \qquad (1)$$

Where, V is a set of vertices (i.e., nodes), E is the set of edges, W: E→ IR+ which could be Euclidean distance, time to travel from one node to the other and the cost (price) to traveling the edge. We represented the accessibility of spatial object in a city constrained by road so we define the dissimilarity between objects by their network distance rather than Euclidian distance. Even if the replacement of Euclidean distances by network distances increases the complexity because, distance between two arbitrary objects cannot be computed in constant time, but an expensive shortest path algorithm is required. The relations between different components of an urban structure are frequently measured along the streets and routes considered as edges of a planar graph, while the traffic ultimate destination points and street junctions are treated as nodes

Network analysis approaches can be useful pointers for a number of interesting urban occurrences by illuminating the significance of even particular junctions in the transportation network so road network helps to find the center point or influential nodes of the network. Influential nodes are road junction which is identified as center node considering distance, time or speed as a weight for computation .Figure 2.4 shows 2 mile network distance coverage from point **A** to point **B** using the Euclidian distance or with the consideration of direct reachability from point A to point B without considering obstacle and 7.5 miles using network distance with the consideration of obstacle and facilitator in real road network. So we have to consider the real problems and opportunities for the movement or spatial connectivity of objects especially in urban area.

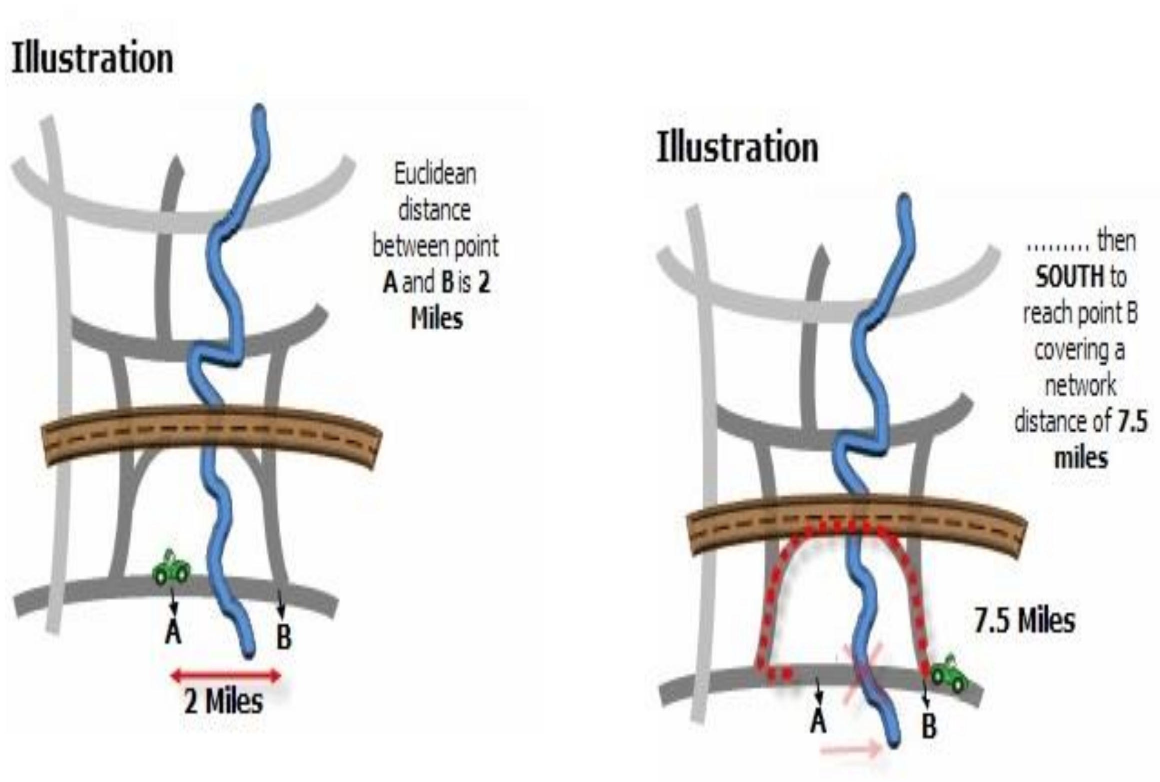

Figure 2.4 : Euclidean vs. Network Distance (Source: www.desktop.arcgis.com)

2.12 Network Centrality Measures

Network centrality measures are the scientific methods of computing the significance of each node in a graph. It tray to find how each graph node is located with respect to the surrounding elements and its graph centrality metrics are equivalent to spatial accessibility measures, but applied on network distance rather than Euclidian [29]. According to the work in **[**29**]** network centrality measures are mathematical methods of quantifying the importance each node in a graph; – Reach, Gravity, Index, Betweennes, Closeness, and Straightness. The closeness measure indicates how close each of these locations is to all other surrounding locations within a given distance threshold which can show similarity is based on the distance between members of the same cluster. For example the work in [26] defined the closeness centrality $Closeness\ [i]$, of a building i in a graph G is the inverse of the total distance from i to all other buildings that are reachable in G within radius r along shortest paths. See Equation (2) below

$$Closeness^r[i] = \frac{1}{\sum_{j \in G-\{i\}, d[i,j] \leq r} (d[i,j] \cdot W[j])} \qquad (2)$$

Where d[i,j] is the shortest distance from point i to destination point j and W[j] is the numeric attribute value given for the destination point j it shows the size, the number of employees they represent or the number of residents they represent. It is one in our case because we compute the distance with the consideration of equal weight for each point.

Closeness centrality shows similarity of objects in the same cluster based on the distance they have in between the above mentioned different mathematical centrality measures.

2.13 Addis Ababa Health Facilities

Based on our case study, service facilities in Addis Ababa (health facilities) can be categorized as hospital, Health center, Health post, medium clinic, primary clinic and specialty center based on the service it provides, the quality services and the number of professionals.

Hospitals general hospital is a medical facility that provides service care for both inpatients and out-patients and treats many types of diseases with professionals. In Ethiopia a general hospital is supposed to serve 50,000 people and provide all types of clinical service including surgery. The market study stated in [30] shows that in Addis Ababa the projected demand of starting from 2006 with 3 additional hospitals up to 16 general hospitals in the year 2020 as shown in Table 2.2.

Table 2.2: The Projected demand of hospital for Addis Ababa

Year	**Total Additional Hospitals (No.)**
2006	3
2007	4
2008	5
2009	6
2010	7
2011	8
2012	9
2013	9
2014	10
2015	11
2016	12
2017	13
2018	14
2019	15
2020	16

The work in [31] defined General hospital site selection requirement as:

- The entry points to the hospital shall be clearly defined from all major exterior circulation modes (roadways, bus stops, vehicle parking)
- Boundaries of the hospital between public and private areas shall be well marked and clearly distinguished. And clearly visible and understandable signage and visual land marks for orientation shall be provided
- The general hospital shall be located away from unordinary conditions of undue noises, smoke, dust or foul odors, and shall not be located adjacent to railroads,

freight yards, grinding mills, chemical industries, gas depot and waste disposal sites.

- The locations of a hospital shall comply with all national and state level regulations applicable to health facilities.
- The site selection criteria shall consider or include the followings, but not limited to:
    1. The minimum size of a general hospital premises shall be 12, 000 -18,000 m2 with two side adjacent road access.
    2. The hospital shall be built preferably in a terrain with a gentle slop
    3. The foundation schemes, soil test and investigation shall be done and it shall comply with the national building code and seismic requirements.
    4. The hospital shall be provided with road access, water supply, electric city and communication facilities.
    5. The building shall be parallel to the wind direction, sun glare and heat. In case difficulties to fulfill these, there shall be technical solutions for such natural effects.
    6. The surroundings of the hospital shall be free from dangers of flooding, landslide, theft, intrusion of stray/wild animals, pollution of any kind (example air, water and sound) and health hazards.
    7. The hospital shall be landscaped, therapeutic, appealing scenery, attractive with green areas/beautiful trees and possible outdoor recreation facilities.

CHAPTER THREE

RELATED WORKS

In this Chapter, the main works related to the proposed model are reviewed. Spatial data mining, GIS-based site selection strategies and hospital site selection analysis are discussed.

3.1 Spatial Data Mining for Facility Location Selection

The process of site selection typically involves two main phases. The first is screening in which the identification of a limited number of candidate sites from a broad geographical area given a range of selection factors. The second is, evaluation in-depth of alternatives to determine the most suitable site [9]. Spatial data mining method is proposed to solve site selection of the service center and field service [32]. City planners are interested in the best feasible way of allocating service facilities (hospitals, fire stations, etc.) to existing and new residential areas. This decision is made according to the local population and constraints [24]. Based on the concept of all demanding point have to be covered and served by the facilities and cost minimization with the consideration of geographical obstacle Javadi & Shahrani stated in [24] three new distance functions. The first function was based on the analysis of shortest path in linear network, which was called shortest path distance (SPD) function. The other two functions, namely path distance (PD) and path to distance (P2D), were based on the algorithms that deal with robot geometry and route-based robot navigation in the presence of obstacles. Because of the distance between the demand points and facilities becoming more realistic in the proposed functions, results indicated the desired quality of the proposed models in terms of quality of allocating points to centers and logistic cost. So the Authors suggested that P2D distance function as an appropriate choice for optimal urban facility location using spatial clustering considering geographical obstacles because of its optimal cost for both logistic and execution time. Zarnani et al. in [33] Faced the same problem to decide the establishment and how many facilities in which locations to ensure the minimization of the overall travel distance of all the customers in a city.

This cost is generally measured by the total amount of distance from the location of each customer request point to its nearest facilities, the authors chosen to partitioning algorithm as the relatively best solution in the context of, " the closer the facilities are to the customers, the sooner and cheaper will be the service provision transaction". Consequently, they study the application of effective spatial clustering methods as solutions for finding the best spatial location of facilities in field-based services in addition to the comparison between the existing clustering. They proposed an algorithm called Fac-means algorithm for efficient integration of the search for the optimal number of facilities along with their optimal location. This algorithm employs the trade-off between the two values in the logistic cost and cost needed for facility establishment in that specific area in a straightforward manner. Fac-means starts by finding the best locations for the establishment of an initial number of facilities and continues by adding new facilities in the most appropriate location if their establishment decreases the customers covering the cost. For each iteration the algorithm tries to further improve the locations by applying K-means. This way, they will be able to search for the optimal number of the facilities beside their optimal locations in a parallel mode.

Traditional clustering algorithms assume that two spatial entities are directly reachable and use a variety of straight-line distance metrics to measure the degree of similarity between spatial entities. However, physical barriers often exist in the real region. If these obstacles and facilitators are not considered during the clustering process, the clustering results are often not realistic. Sun et al. [34]. Proposed the artificial immune clustering with obstacle entity (AICOE) algorithm for clustering spatial point data (the spatial distribution of communities) in the presence of obstacles and facilitators; in the consideration of distance measure between sample points in object space is an important component of a spatial clustering algorithm. These authors try to show the result of different scenarios like when the rivers and hill as obstacles are not considered, if the obstacles are taken into account and bridges as facilitators are not considered and both the obstacles and facilitators taken into account.

As a result, they proposed an algorithm that handles linear obstacles and planar obstacles like rivers and lakes respectively.

When traversing linear obstacles, facilitators are also taken into account for path construction and the algorithm constructs approximate optimal path for linear obstacle. When traversing planar obstacles, path is generated by the method to construct the minimum convex hull by Graham algorithm. So they tried to provide relatively closer solution for the existing on ground problems by considering spatial constraints and spatial object connectivity facilitators and replacing Euclidean distance by obstructed distance.

3.2 GIS Based Multi Criteria-Analysis for Hospital Site Selection

Service facility site selection related to several aspects of the society with mixed views and debates on which criteria are most important would confuse even the experts. To reduce these complex problems with multiple criteria into finest ranking of the best scenarios from which an option is selected; Zhou and Wu in [6] used a GIS-based MCA together with necessity tests and sensitivity tests, which multiplies conflicting criteria that are essential to be evaluated in decision-making. The criteria are weighted according to their importance, and their weights have a more or less favorable to the final decision than another. GIS-based MCA model resolves the factor criteria and constraint criteria separately and integrates them together by multiplying suitability to the objective being considered with integrated constraint and gets the final output, instead of taking all the criteria in the MCA process at the same time, the authors perform necessity tests and sensitivity tests on the factor criteria. The necessity tests check the necessity of factor criteria. Adding each factor criteria with some other factor criteria one by one if the variation changes a lot, the added factor criterion is necessary, otherwise is not. The sensitivity tests assess the sensitivity of the result to the weight change of factor criteria. Both tests make the decision makers be aware of which criteria are really necessary and how the results are sensitive to the weights change in order to overcome the drawbacks of too many repeated operations.

After that, the candidate sites with maximum values are selected. Finally, by taking into account the sites' surrounding conditions and subjective parameters, the optimal site to the new site will be decided.

3.3 Hospital Site Selection Analysis

Hospital site selection like Michigan Community hospitals, which considers the geographical and historical factors including the distribution of population, is becoming the main concern for successful and sustainable site selection. The work in [4] combines the concept of distance for existing hospital and road network density to estimate travel time.

The study is unique and relied on well accepted theoretical and computational foundations for support. The authors set the following basic requirements.

- 3 mile spatial resolution, after a major experimentation the models were recreated to run in 1-kilometer cells and results using the 1-kilometer cells are presented to compute travel time over space because the initial consideration of the 3 mile spatial resolution was much larger for the selected study area.
- All places in the state must be measured because the process is repeated until all cells on the grid have been assigned an accumulative cost. The cell location with the least cost is then removed from the list. Finally, the least-accumulative cost to each of the neighbors of the cell that was just removed from the list is determined.
- 30 minute travel time maximum to suitable hospitals. The state was concerned with identifying populations with long drive times to existing community hospitals so 30 minutes travel time from existing hospital became the maximum travel time to identify accessible and inaccessible of the study area. “Areas falling below a particular time threshold would be identified as relatively inaccessible”
- Variations in road types must be considered. Travel time maps of the study area are mainly based on three major road types (Arterial, Collector, and Local) because travel time or speed set for each road by the authority is depend on road type and in addition, road further grouped in rural and urban road because urban roads will have delay during rush hour and the authors used 25% reduction in speed limits.

The cost grid, or travel time is based on average representative speeds of three major road types (Arterial, Collector, and Local) and roads are further divided into urban and rural and speed limits defined by this road types.

The work in [4] clearly stated that, Euclidean distance function class fails to effectively capture the variation in transportation networks so the authors used “Weighted distance functions” and, ultimately, “Pathdistance” selected for the travel time methodology to determine the shortest weighted distance calculation of travel times.

After identifying the size and the cost of each grid cell the cumulated cost of the source cell will be identified and in order to identify all areas which are beyond the determined threshold based on the minimum speed found in the cell. The overlay of the output map for travel time from selected hospital, travel time by 25% existing selected hospital for urban speed reduction, selected area with population greater than population 50,000 and wholly outside of the 30-minute travel time limit two areas identified.

3.4 Summary

In [24], the author's proposed three different algorithms and selected the one which is best for both logistic cost and execution time only depending on the performance of the algorithm. The work in [33] considers total establishment cost, logistics cost and response time as the main criteria that contribute to their work. The authors also proposed a new algorithm that will identify the number of facilities with their optimal location in association with k-mean algorithm. But their work doesn't consider any kind of obstacle which makes the result unrealistic.

The work in [34] proposed an algorithm that mimics the natural biological function of the immune system and also considers the existing defined obstacles. The comparison between their proposed algorithm and previously existing algorithms shows the new algorithm is more effective, even if the spatial connectivity of the clustered object is not considered at all. All the above algorithms use Euclidian distance which is not realistic, especially in urban areas because, in urban area any spatial object connectivity including the movement of human being should be defined with the consideration of obstruction.

The authors of [6] combined spatial and non-spatial data to construct visualized information that can be easily understood and analyzed by decision makers. Conflicting criteria and parameters can also be re-checked and evaluated based on necessity and sensitivity test. But the work focuses on the analysis of specific areas (candidate locations) in depth which could come after finding the general location using some other techniques.

The work in [4] focused mainly on the spatial distribution and density of the society relative to their driving time to the nearest existing selected hospitals. The authors proposed "pathdistance" function to capture the variation in landscape and determine the shortest weighted distance or cumulated travel cost from each point to the nearest hospitals using the existing road network. Road type variation in speed and urban road network, traffic jam are also considered, area out of the specified travel time (30 minutes) is identified as inaccessible and inaccessible area with population density more than 50, 0000 will be the selected area for newly established hospitals.

We can identify two groups of theories used for handling the above discussed problem. Based on the first group, researchers pay attention only into the performance, scalability and efficiency of algorithms used to identify optimal site selection using different spatial data mining methods. Based on the second group, researchers also pay attention into handling different conflicting ideas and try to fulfill the needs and requirements of different interested groups. Thus, they work on selection criteria that satisfy social and environmental requirements, which are enforced by legislations and government regulations in addition to the preferred algorithms.

In Addis Ababa starting from its foundation proper facilities site selection was not practiced and harmonized with urban planning so, most facilities found in the city assigned on existing buildings which are not built for that specific service or site selection for the specific service was not totally considered its accessibility and suitability for the surrounding residence except the existence of free and cheap land, as a result, the above discussed groups of theories are not practiced, even if many minimum standards for site selection criteria are exist for different facility types**.** Our work depends on the two groups of theories above and uses the output of spatial data mining called group analysis and suitability analysis as an input to leverage the positive outcome of the two approaches for service facility site selection to fulfill the both concerns at the same time.

CHAPTER FOUR

MODELING SERVICE FACILITY SITE SELECTION FOR ADDIS ABABA CITY

In urban location selection, finding the optimum location that increases service reachability with the consideration of environmental, legal, political, technical aspect from many interdisciplinary ideas is the main problem. Urban society uses road and roadsides to move from one place to the other by its own foot or uses vehicles. Network analysis based on the existing road helps us to conceptualize the spatial relationship of urban objects with their obstructed distance because the spatial connectivity of urban objects depends on the route used by the residence. We can extract and identify different data which will use as an input for spatially enabled optimum location selection from existing road network. Additionally, the consideration of spatial connectivity within urban objects using the road will also help us to define the existing obstacles and facilitators for the accessibility of the facility. Urban network analysis tool, spatial statistics tool, and network dataset tool are used for data preparation, extraction and analysis. Open street map is our source of road data, but after overlapping and analyzing the downloaded data with the current satellite image the result showed clearly the quality gap and the next work we have to do with data pre-processing section.

From the data, we found some road edges shifted from its exact position, some edges bypass or underpass to each other, some edge junction points don't intersect at all and created Network Island. Some existing road edges are missed; some existing road edges are not currently exists because of a design change or some construction. Some edge which should be merged is split and vice versa, some two way roads represented as one way and some attribute values of edges are missing. So, our data pre-processing activities fulfill the above gaps, one by one, because, the result of spatial network analysis and clustering will be totally wrong and site selection based on wrong data will lead us to wrong decisions. We analyzed site selection problems in urban areas for service facility.

The proposed model uses GIS based analysis methods using a hybrid model of knowledge discovery called spatial clustering analysis and Multi criteria analysis using AHP method to leverage their advantage at a time that makes our model different and new for the case of Addis Ababa. We generalized and used the spatial clustering analysis, which uses road junction point and road network, spatial weighted matrix as an input and multi criteria analysis uses different constraints and factor map inputs to site suitability analysis. We presented the general model used to cover the whole process in the Figure 4.1. The main component of our work and their interaction to each other are also described.

Main Components of the Model

Our model composed of the following main components such as, pre-processing of the data, data mining, and evaluation of the discovered knowledge, GIS-based MCA and use of the discovered knowledge as shown in the Figure 4.1. Pre-processing of the data used to prepare the input data used for the analyses. Data mining used to derive the new, novel and usable knowledge from the existing preprocessed data. Evaluation of the discovered knowledge used to check the impact of the discovered knowledge with domain experts and used to revise the entire process which could see alternatives to get the best result from the process. GIS-based MCA analysis is used to identify the suitability of an area based on selected criteria. Use of the discovered knowledge used to plan where and how to apply the discovered knowledge based on some applications.

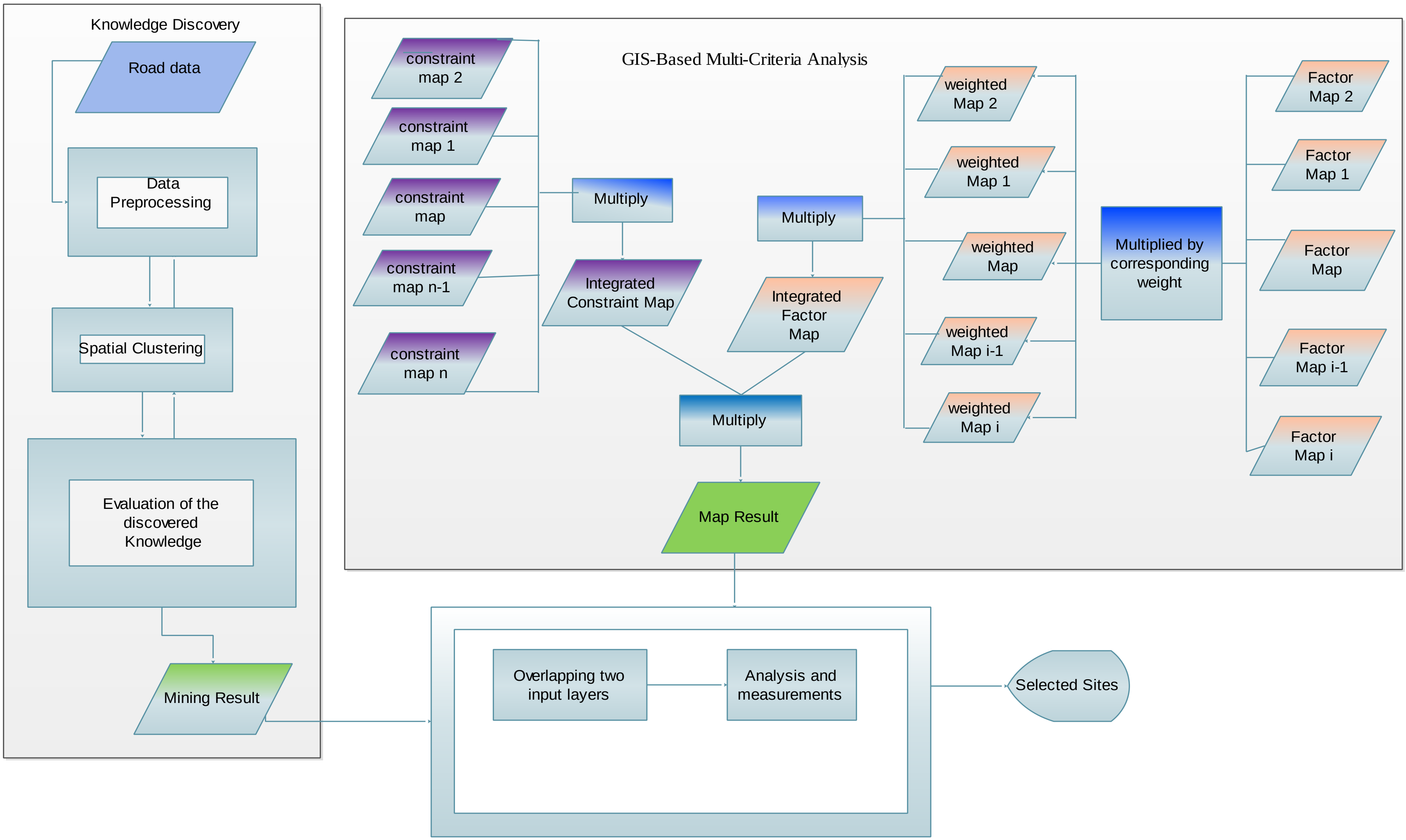

Figure 4.1 : Architecture of Service Facility Site Selection Model

4.1 Data Pre-processing

Spatial data pre-processing is time consuming and higher effort task [35]. This step concerns deciding which data will be used as input to the data mining process in the subsequent step. In spatial data pre-processing the work in [35] stated that spatial data pre-processing involves spatial outliers in point data, spatial outliers in polygon data, pre-processing for unified format and projection, missed values and features, projection, conversion of spatial value at different granularities (point to polygon) and spatial granularity dealing with multiple formats. In this work our input data for knowledge discovery are spatial point data and spatial weighted matrix prepared from road network dataset thus our data preparation includes:

- Projection of references.
- Adjusting road type and shifting from its position.
- Identify and remove spatial outlier in road data outside boundary.
- Digitizing missed road features and fulfilling the missed values.
- Merging and splitting road edges.
- Snapping road junction points to removing island.
- Preparation of road network dataset.
- Extracting of road junction points and adding attributes to be used as a spatial analysis field.
- Generating network spatial weight matrix.

In spatial data preprocessing, the first thing we have to consider is the spatial reference of all layers. It should have the same reference, so we have to project all our layers in a projected format with similar reference. After formatting our data into the same reference, we have to adjust the road type and siftings from its normal position because, some spatial layers with the same projection and reference may shift from the normal because of digitization quality and some edits made to the layers. Depend on the data size we have and our study area, some segments of layers or points could be out of scope or out of boundary so we have to remove those layers. After identifying the required layers, we have to digitize the missed layers and fulfill attribute values. The spatial connectivity of the road edges could be out of rule so, in order to keep road data quality standard and fully answer the requirements of spatial analysis on a road network, we have to merge the road that should be merged and split the road that should be split.

Snapping some road junction points for the identified road network island to complete our road network analysis without error is also mandatory. Preparation of the road network dataset is the beginning to qualify the previous data pre-processing steps, and road dataset will be created if all steps are prefect. After successful network dataset preparation, we have to extract and prepare road junction points and network spatial matrix. All the above processes are accomplished using different ArcGIS tools and Section 4.2 shows each detail implementation.

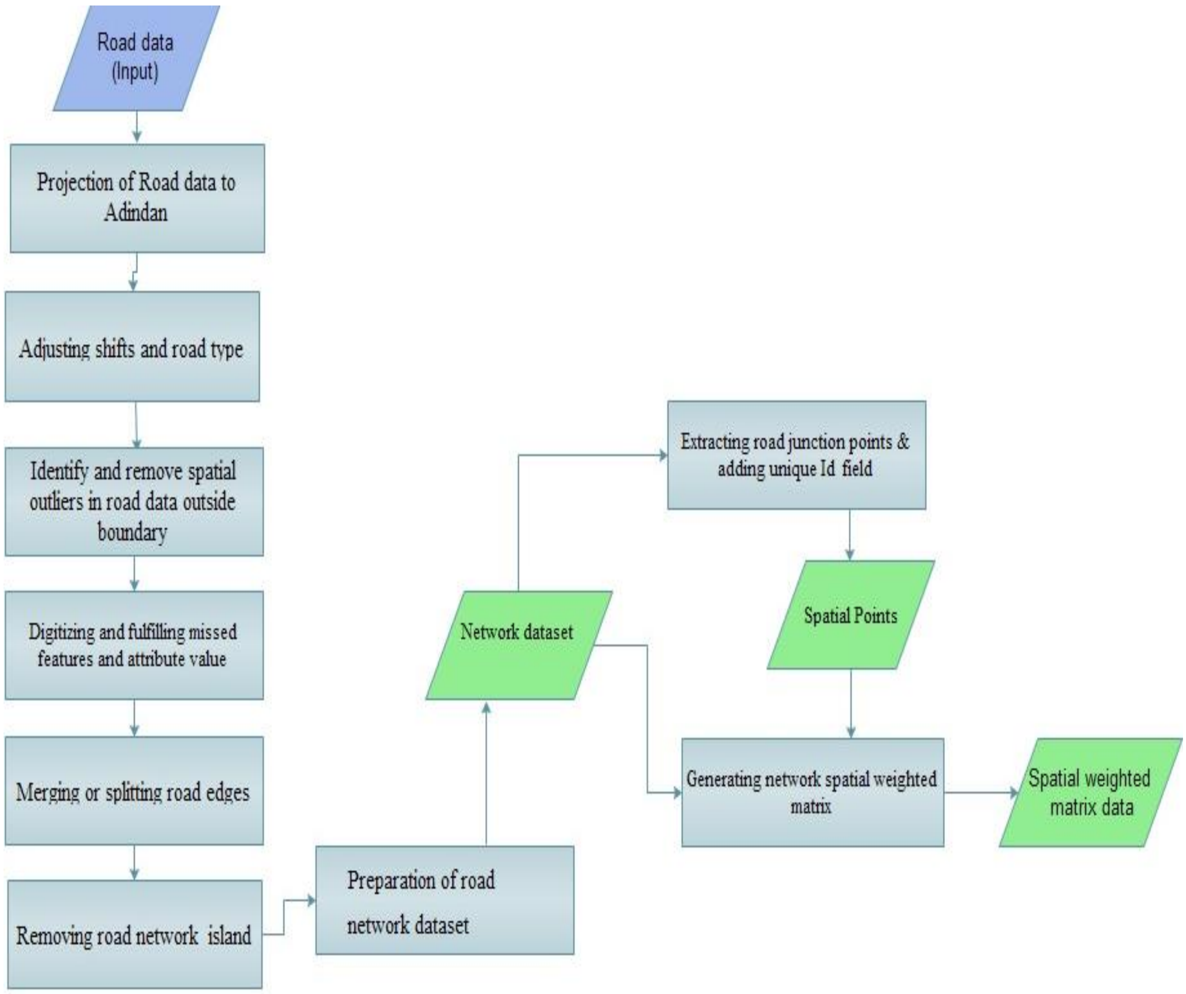

Figure 4.2 : Data Pre-processing Steps for Spatial Data Mining

- **Projection and Transformation**

Data projection and transformation of spatial reference from one to the other has three reasons: the first one is if we have more than one data layers which have different reference, in order to overlap and see the topological patterns of those layers without shifting, we have to transform into the same reference. The second depends on the type of spatial analysis we perform. In our case, we have to Project our spatial data reference from GCS_WGS_1894 in to Adindan-UTM-Zone-37N because Adindan reference is the closest reference exist close to our country to maintain and get accurate earth curvature during projection for large scale data or for detail work like city data analysis. The third, our analysis need projection from geographic coordinates to UTM to use planner units like meter. The downloaded, open street map has WGS reference so we projected our imagery and vector data into Adindan_UTM_Zone_37N.

- **Spatial Outlier of Road Data**

We overlapped Addis Ababa boundary with road data to view the road out of the boundary and the data that was out of bounds has been clipped because our work is for in the case of Addis Ababa.

- **Digitize and Add Road Features**

The existence of road edges that are missed, forced us to digitize and fulfill the missed edges of the image. So we carefully examined the missed road segments and digitized manually using Arc Map 10.4.1 and removed those road segments not exist currently. The digitization process follows the road digitization standard and filling all needed attribute information.

- **Merging and Splitting Road Edges**

Some parts of road edges didn’t segmented properly, road edges which have junctions should be split and a single edge without junctions shouldn’t be segmented. So we merged and split the existing road data properly as needed because, for example in road round if we don’t segment the circle representing the round with intersecting edges from different directions our route analysis will go round 360^0 around the roundabout and can’t continue further.

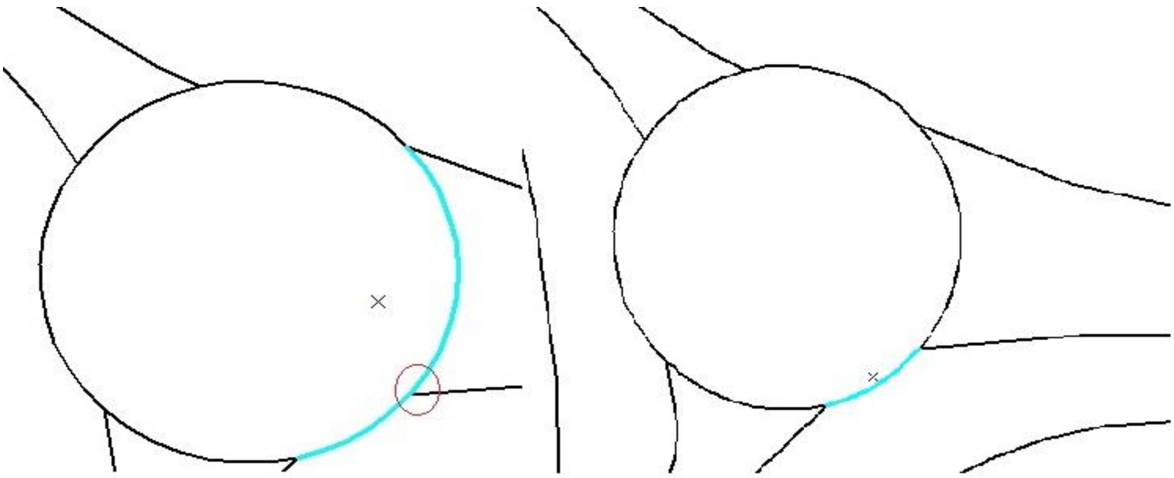

Figure 4.3 : Road junction need to be split at junction point vs. splinted junction

- **Removing Island and Snapping Road Junction Points**

When one area has a single edge to connect with other network segments and that road junction point didn't intersect with the other road network during digitization or road data preparation, a road island will be created. So the area will not be reachable at all and will be out of analysis because of its connectivity thus, we removed existing road network island by properly merging the road edge junctions of the existing road.

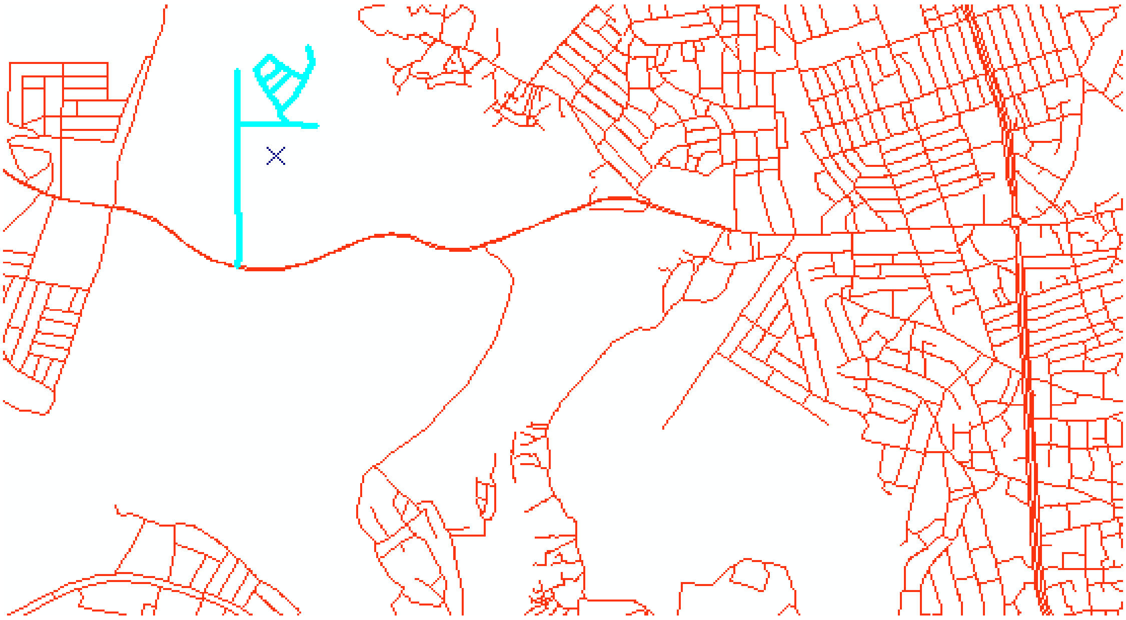

Figure 4.4 : Road with an Identified Island

- **Adjusting Road Type and Shifting**

Some road types may be changed during road reconstruction and must be adjusted based on the current up-to-date image data we have and shifting of the actual image also adjusted too, but the spatial reference of both the road and image data should be the same so we adjusted those road segments with different type and those who have shifted from the current image.

- **Fulfilling the Missed Attribute Values**

Attribute values like road name, direction, distance, speed and some other related filed values have been filled if they are missed. Sometimes during splitting road edge, we have been calculating the new split distance values otherwise, the previous distance value for both splits will remain in the field is wrong.

- **Creating Road Network Dataset**

Road dataset creation is one part of data preparation for spatial data mining, which defines the road as spatial weighted graph and the graph contains spatial points as a junction and edges with their weight. We created network dataset using network analysis tools for the analysis and we extract junction points as spatial points uses for clustering network spatial weight matrix to define the spatial connectivity of urban objects like road junction and adjacent parcels used to define spatial constraints during group analysis and road edge attributes used as a parameter for spatial clustering by considering the route distance or time.

- **Extracting Road Junction Points**

Junction points are intersection points of road edges which are spatially enabled or geo-referenced. We used road junctions, as communities live in the surrounding environment that can spatially be clustered based on their spatial distribution because any parcels or blocks which residence lives in it are between these road junctions so we can represent the spatial connectivity of community by road junction points and weighted edge. Spatial connectivity of the community is represented by road because any movement of the residents is using road or road side, so road connectivity easily represent the spatial connectivity of the resident .When we create network datasets, junction points will be created separately as a single layer. We added one attribute which its values uniquely identify each feature and we prepared it for group analysis or spatial point clustering purpose. Figure 4.5 shows sample junction points we extracted for ‘Kirkose’ sub city as a sample.

Figure 4.5 : Sample Road Junction Points

- **Generating Network Spatial Weight Matrix**

Spatial weighted matrix is a matrix that created in the format of little Indian and ASCII character; it conceptualizes the spatial relationship of creating point data based on network dataset and its selected attributes as spatial weight. Group analysis uses spatial weighted matrix as a spatial constraint parameter in order for the system to use a minimum spanning tree with K-Mean algorithm because the distance between two points should be based on route distance. We generated Network spatial weight matrix from the extracted road junctions and network dataset and define time as a spatial weight of the spatial relationship, because time depends on the speed and distance of the road which represent the road weight

Distance is computed automatically during road extraction and speed depends on road type which is constant value. Figure 4.6 Shows sample road data attributes used for matrix computation

OBJECTID *	FT_Minutes	TF_Minutes	Meters	Speed__Lim	Length	Road_Class	Shape *
136	0.01826	0.01826	18.260292	60	18.260292	1	Polyline
137	0.049972	0.049972	33.314485	40	33.314485	2	Polyline
138	0.043105	0.043105	28.736789	40	28.736789	2	Polyline
139	0.07315	0.07315	48.766528	40	48.766528	2	Polyline
140	0.047766	0.047766	31.844174	40	31.844174	2	Polyline
141	0.05214	0.05214	34.760284	40	34.760284	2	Polyline
142	0.065219	0.065219	43.479585	40	43.479585	2	Polyline
143	0.013913	0.013913	9.275091	40	9.275091	2	Polyline
144	0.035278	0.035278	23.518541	40	23.518541	2	Polyline
145	0.0483	0.0483	32.19997	40	32.19997	2	Polyline
146	0.041889	0.041889	27.925955	40	27.925955	2	Polyline
2	0.125708	0.125708	125.70828	60	125.70828	1	Polyline
65	0.133289	0.133289	133.289418	60	133.289418	1	Polyline

Figure 4.6: Sample road attributes

We can convert spatial matrix file (.swm) to a table which can be understandable by human using spatial statistic tool. Figure 4.7 shows the converted sample table

OBJECTID *	UID	NID	WEIGHT
103662	3456	3892	0.035104
103663	3456	3639	0.035104
103664	3456	3936	0.035104
103665	3456	3898	0.035104
103666	3456	3900	0.034308
103667	3456	3910	0.034039
103668	3456	3913	0.033731
103669	3456	3908	0.033704
103670	3456	3907	0.032495
103671	3456	3923	0.031697
103672	3456	3944	0.031137
103673	3456	3927	0.030826
103674	3456	4136	0.030592
103675	3456	2932	0.030514
103676	3456	2946	0.030499
103677	3456	4151	0.030251
103678	3456	3931	0.03008
103679	3456	2906	0.02991
103680	3456	2913	0.029658
103681	3457	3439	0.033333
103682	3457	3471	0.033333

Figure 4.7 : Spatial Weighted Matrix Converted into Table Form

4.2 Spatial Point Clustering

We used road junctions for group analysis to cluster the community and the parcels found in between using community movement. The movement of community other than air flight is based on the road so, we used road junction as a community and road edge as a spatial relation of the community. We used ArcGIS spatial statics tool for spatial data mining method called spatial point clustering. The parameters we used for the analysis are spatial point data, spatial weight matrix and some other parameters provided by the tools. Figure 4.8 shows group analysis result of Nifasilik_Lafto sub city. The number of clusters is two or k=2.

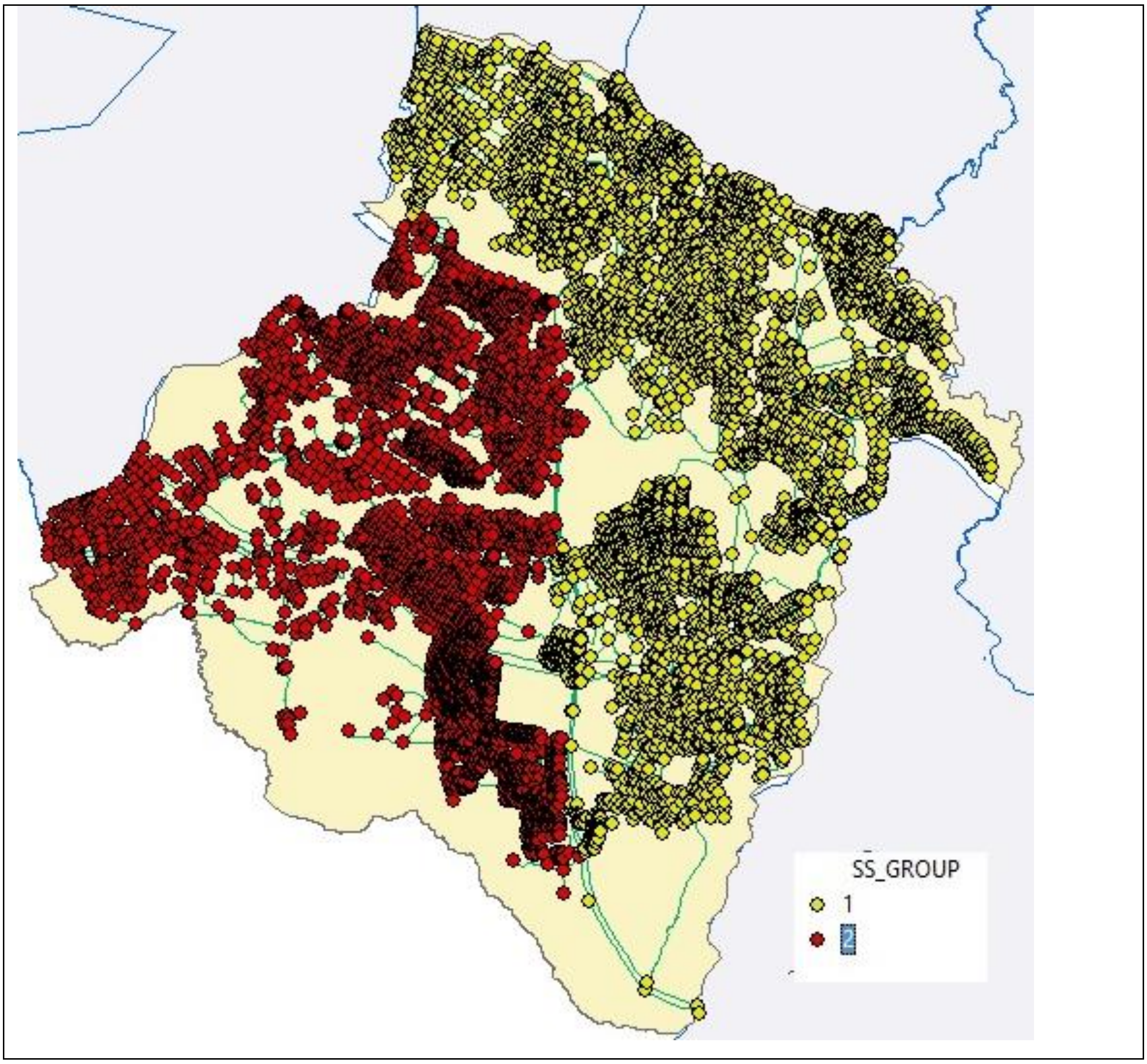

Figure 4.8 : Sample Output for Group Analysis

4.3 Spatial Data Mining

Mining of spatial data helps us to understand the inter-relation between different spatial entities which are depended on each other and affect the changes at one location to other. After the preparation of the needed data to mine, our spatial data mining will be conduct by using ArcGIS tool called group analysis tool, using spatial point data and network spatial weight matrix defined with (.swm) extension as a spatial constraint. The inputs for the spatial data mining are spatial point data used for clustering and network spatial weight matrix as a spatial constraints. We used network spatial weight matrix for two reasons, the first one is to conceptualize the spatial relationship of the input data points and the second is to define constraints and facilitator. The output of the result will be stored in the form of point layer grouped by the number of clusters. Optimal number of cluster can be suggested by the group analysis tool, if we want to compare the input parameter and the number of cluster suggested by the tool based on the input data we gave for clustering. During the clustering process if we use spatial constraints as a parameter the tool use minimum spanning tree algorithm, in our case, we used networked spatial weighted matrix as a spatial constraint so SKATER (Spatial "K"luster Analysis by Tree Edge Removal) method is employed. If not, it use Kmean algorithm.

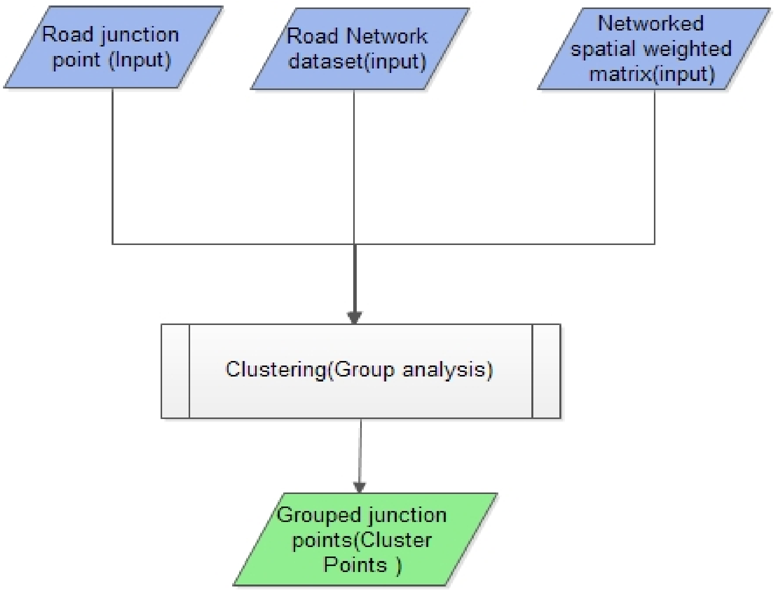

Figure 4.9 : Clustering Analysis

4.4 Evaluation of the Discovered Knowledge

Based on the data mining goal domain experts will check if the discovered knowledge is novel and interesting. The entire process is revisited to identify which alternative actions could have been taken to improve the results. The knowledge discovered from point data clustering is mainly identifying best clusters which their similarity is based on the overall distance between them is minimizing. Evaluation of the result is then depends on the activities and strategies we use during knowledge discovery steps starting from understanding of the problem domain up to data mining steps then the best strategies and model will remain intact. In the case of our work evaluation of the discovered using report file generated during clustering process by using variable wise summery and F-mean statistics for each number of cluster.

4.5 GIS Based Multi Criteria Analysis (GIS-Based MCA)

MCA largely use for resource allocation and land suitability analysis. It is a procedure that typically multiplies conflicting criteria that are essential to be evaluated in decision-making called factor criteria and constraint criteria. Factor maps are represented as spatial distributions to display the opportunity criteria and the quality of achieving an objective. Constraint maps are limitations or restrictions which prohibit certain elements to be taken into account the analysis. We used IDRISI Selva for this analysis because it is GIS software that includes decision-support modules based on AHP, and some other analysis types, among others, plus it has a wizard to assist in selection of appropriate decision techniques to conduct the following activities

After pre-processing has been done on any factor criteria converting the data to raster, reclassifying, stretching processes are conducted to make the layers ready for the MCA process called Factor map [6]. Raster data are needed to execute the MCA model using IDRISI, Shape file format files are thus converted to raster format. In the MCA model we treat factor and constraint criteria independently as shown in Figure 4.5. Constraint maps will multiply each other and weighted maps which is derived from Factor map and its corresponding weight will also multiply each other, then the integrated constraint and factor maps will produce the final map result as an output from the MCA model using the AHP method which will use as one component of site selection.

4.5.1 Data Collection and Preparation for GIS- Based Multi-Criteria Suitability Analysis

We used mainly open street map, United States geological survey (USGS) and Google maps for both factor and constraint criteria used for suitability analysis. In our case we downloaded most of the layers from open street map and for digitizing and careful inspection of the downloaded data, we used imagery from Google earth additionally we collected DEM data from USGS. For constraint criteria mentioned above, we digitized university areas, airport, west disposal and green belt using satellite image and collected DEM data from USGS and generated slop data. Ethiopian Standard Agency tried to identify factors and constraint criteria for health facility site selection, but factor setting for each criterion is not specifically set as a rule. So we used the best practice of the work in [6], in addition, we also used factor criteria settings in our work .In GIS based multi-criteria suitability analysis, we used different techniques for data collection and preparations for the analysis depend on our objective. In our case we used the following procedure to prepare factor and constraint map as shown in Figure 4.10

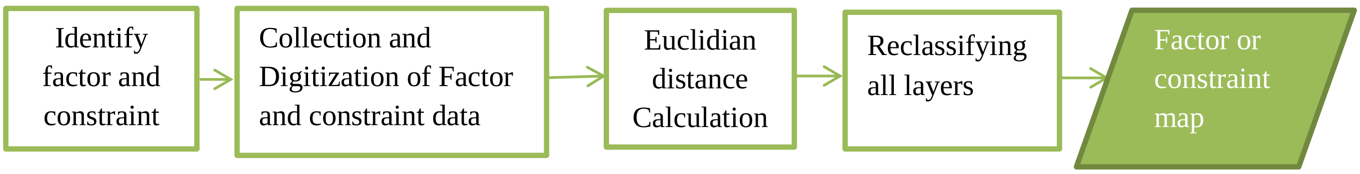

Figure 4.10 : Factor and Constraint Map Preparation Steps

4.5.2 Identify Factor Criteria

We adopted all identified factor criteria, constraint criteria and factor criteria settings except west disposal, are based on best practice on [6]. West disposal is added because it shouldn't be compromise and public toilet is rejected considering the current condition of Addis Ababa.

- Existing hospitals.

  For fair allocation of resource and having a fair market share of the existing and newly built service facility, we have to seriously consider the existing facilities in our analysis and keeping a fair distance between them. Even if the Ethiopian government didn't clearly stated the distance between hospitals or service facilities, in this study, we wish to keep the distance 500 meters away, the further away better to serve distant residence.

- Main roads

Road play a major role for the reachability of any facilities anywhere, especially in urban areas. The Ethiopian government has also stated clearly on the service site selection standard as "the selected site should be adjacent to the road" [7]even if the type of road which should be adjacent is not stated. Some roads like ring roads and highway road have noise from motor vehicles passing by influences the patients in the hospital. Therefore, buffering the road in 100 meter radius is mandatory. The nearer outside the buffer is, the better.

- Sub_roads

  Sub roads are roads which consider being relatively good in noise disturbance and adjacent to the selected site will increase the site accessibility. The nearer away from the sub road is, the better.

- Railway

  Addis Ababa railways are not widely in use like some other cities, but we have to consider the existing railways. So 200 meters far away is the better.

- River

  In some other cities, service facilities should away from the river because drainage discharge and some other factors could contaminate the river, but the rivers in Addis Ababa are badly contaminated and became a source of bad odor and a health problem by itself so 300 meters far away is the better.

4.5.3 Identification of Constraint Criteria

In our case we selected six constraint criteria which include:

- Altitude

Altitude has a significant effect on the construction of hospitals because flat area will affect the west discharge from the hospital and susceptible for the flood and hilly areas will affect the construction cost, accessibility and residence in hill area will be scattered. So as the standard of Ethiopian government stated the altitude should be gentle slope.

- Airport area

The boundary inside airlines is not suitable for new hospital construction and no need to construct hospitals inside airlines because harsh airplane sounds will disturb the patient and construction inside the airport site will affect both airport and hospital services.

- West disposal area

West disposal areas like ‘koshe’ produce a bad odor and make the surrounding environment polluted for the patient so buffering 1killometer is needed.

- Green Belt

Green belt like parks, green area and forests should be preserved so building hospital inside is not necessary.

- University Area

Educational sector could have their own small service post and constricting hospitals inside the area will affect the teaching learning process so it is not advisable as stated in [6]

4.5.4 Euclidean Distance Calculation

Euclidean distance measure the distance from every raster cell to the nearest source or standardizes the distance based on the criteria given for both identified factor and constraint. For factor criteria, suitability to attain the objective could be increase or decrease when we close to the criteria map depend on the criteria map type. For constraints, we have two options the area inside constraint map is not possible at all and we can consider outside the boundary without any level of importance. Table 4.1 shows the criteria setting for Factor criteria and Table 4.2 shows criteria setting for constraint.

Table 4.1 **:** Factor Criteria Setting

NO	**Factor**	**Setting**
1	Existing Hospitals	The further away from the existing hospital (>=500m), the better
2	Main Road	The nearer away from the road (>=100m), the better
3	Sub Road	The nearer away from the road , the better
4	Railway	Further away from Railways (>=100)
5	River	The further away from the river (>=300m), the better

Table 4.2 **:** Constraint Criteria Setting

NO	**Constraints**	**Setting**
1	Altitude	Between 7% - 15%
2	AirPort	Inside the area = 0 outside = 1
3	West Disposal	1000 meters buffer zone is set. Inside = 0, outside = 1
4	Green Belt	Inside =0 outside = 1
5	University Area	Inside the area = 0 outside = 1

Based on the above given setting Euclidean distance is calculated for all factor and constraint criteria. Figure 4.11 shows the Euclidean distance for existing hospitals and the color grid shows continuous measure of Euclidean distance from closest existing hospitals to each cell up to maximum 11,366 meter.

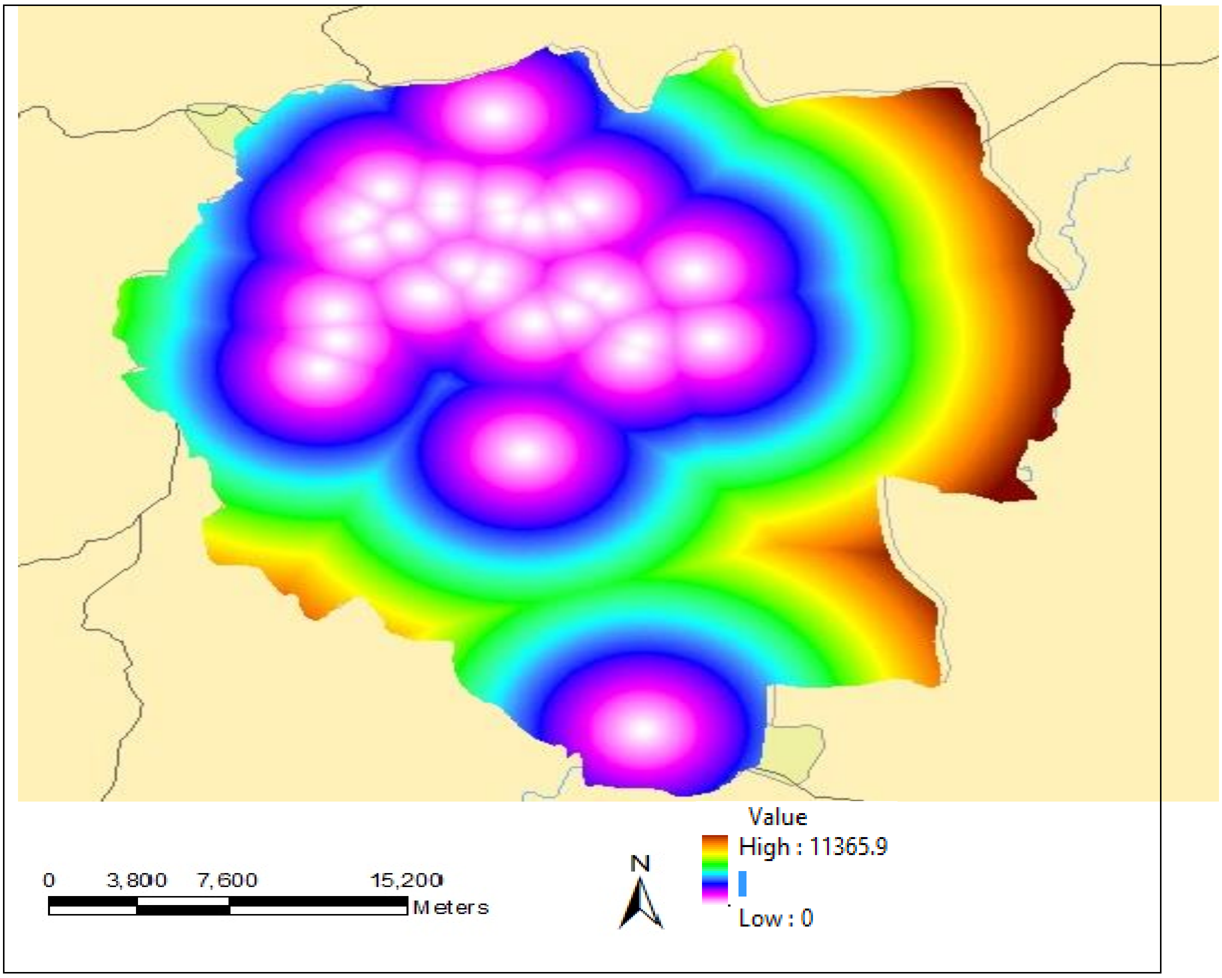

Figure 4.11 : Euclidian Distance for Existing Hospitals

4.5.5 Reclassifying Layers

Reclassified layers will use as the input for MCA analysis for both factor and constraint criteria. It is useful when we want to replace the values in the input raster with new values or when we may want to simplify the information in a raster. Another reason to reclassify is to assign values of preference like in the case of constraints to have only 0 and 1 value, sensitivity, priority, or some similar criteria to a raster as shown in Figure 4.12 .We reclassified the values of the raster into 20 classes for existing hospital using the above Euclidean distance image shown in the Figure 4.11 which means continues values from 0 - 11,366 meter is categorized and given a new value for range of values from 1 up to 20 for each cell. The number of class depends on the distance which is considered for factor criteria setting.

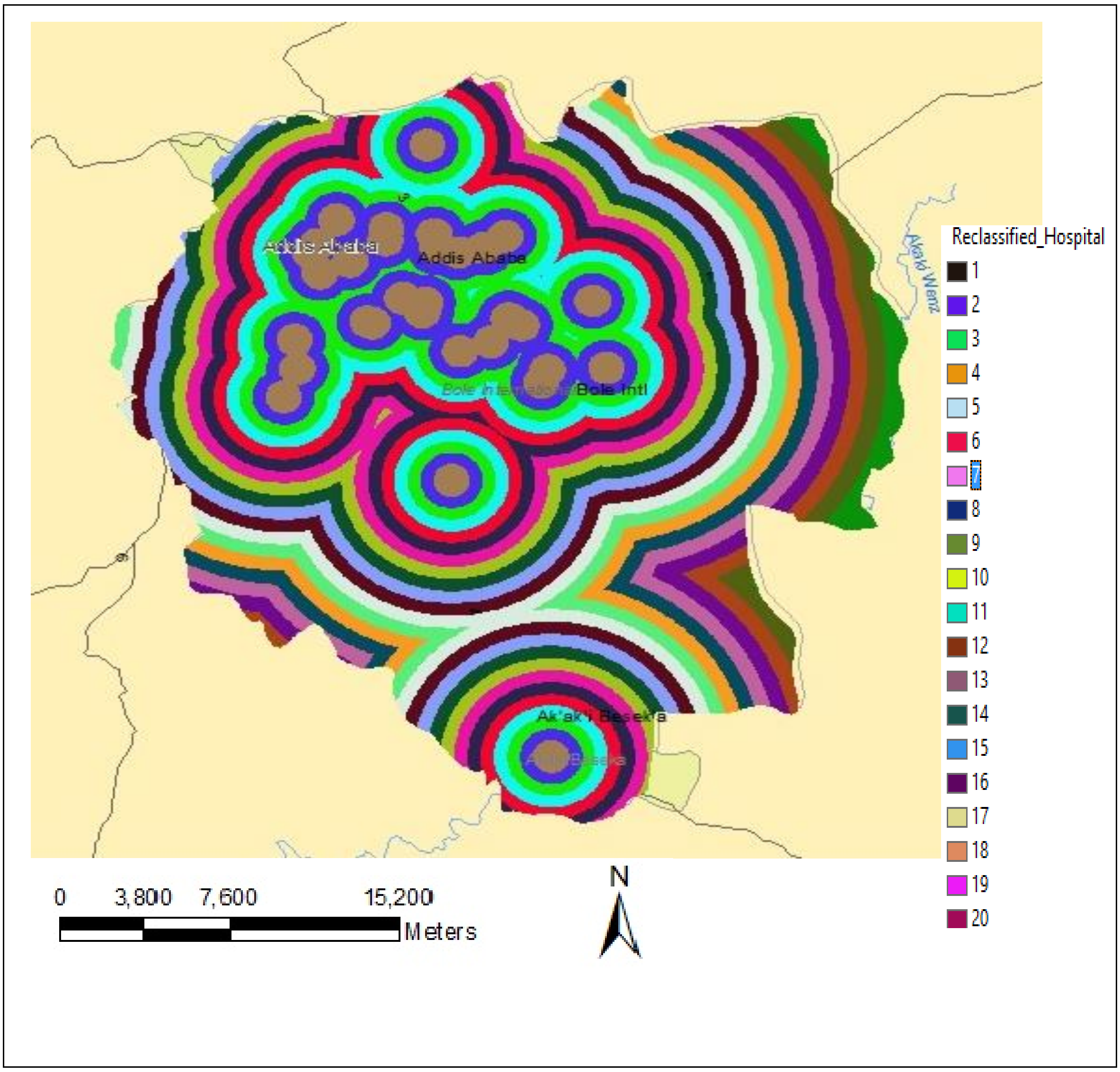

Figure 4.12 : Reclassified Layer of Existing Hospital

We also reclassified all constraints maps with raster cell values with values of 0 and 1 as shown in Figure 4.13. 0 represent restricted area and 1 represent allowed area.

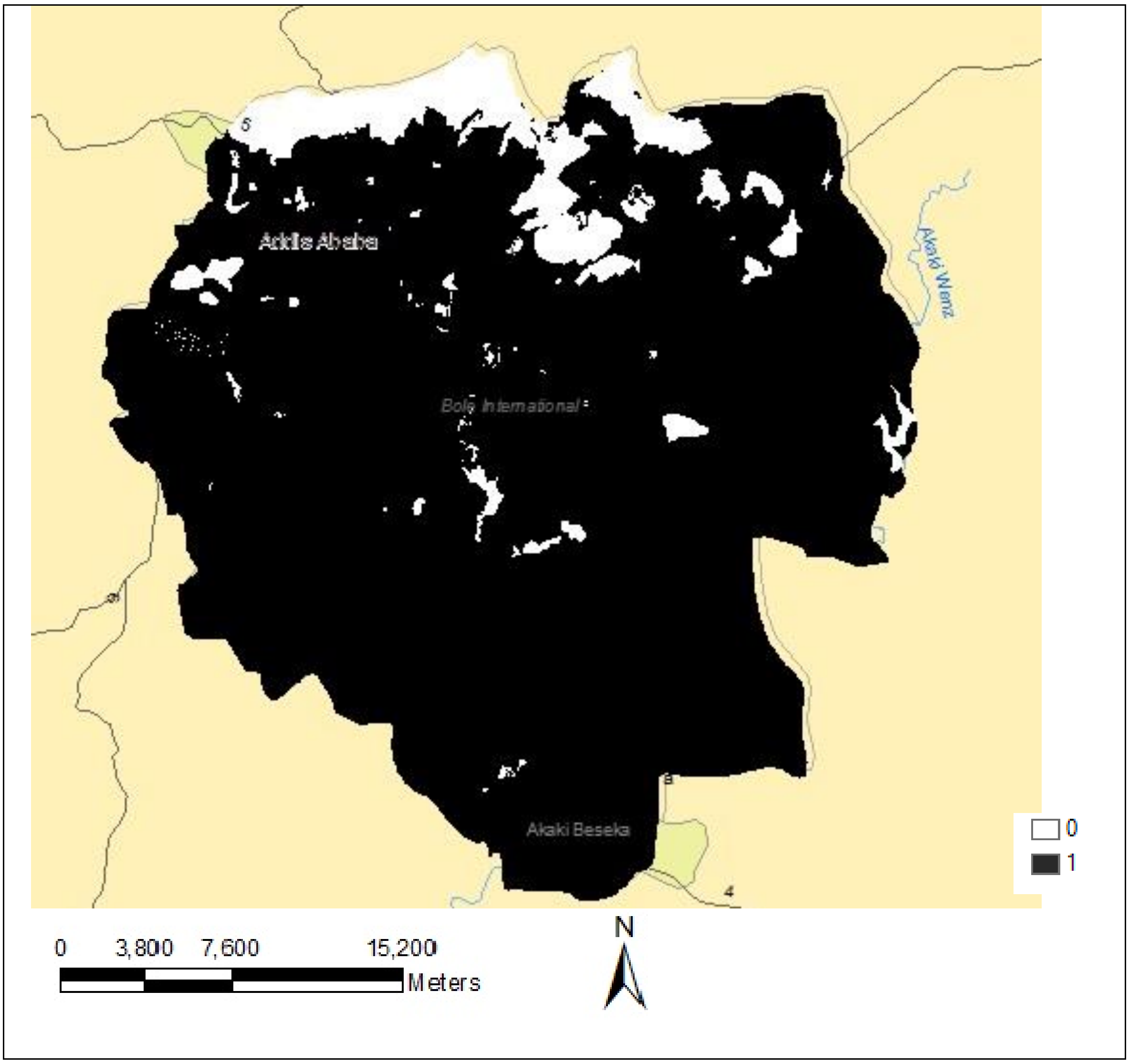

Figure 4.13 : Reclassified Green Belt

4.6 Suitability Analysis

Our suitability analysis uses AHP method, so the weights to the decision making criteria are derived from the pairwise comparisons of the relative importance between two criteria (the sum of the weights equals to 1). Our suitability analysis start by defining and giving the relative importance of our identified factor criteria, Table 4.3 [9] shows intensity importance of factor criteria.

Table 4.3: Scale for Pairwise Comparisons

Intensity of importance	Description
1	Equal importance
3	Moderate importance
5	Strong or essential importance
7	Very strong or demonstrated importance
9	Extreme importance
2,4,6,8	Intermediate values
Reciprocal	Values for inverse comparison

(Source: Risk and Policy Analysts Ltd (RPA) in [17])

After factor map is produced each factor criteria map will multiply by the comparative weight given for each factor criteria. The AHP weight setting is then handle by IDRISI Selva decision support tool after we gave factor criteria map and intensity of importance for each criteria the tool will calculate the AHP weight and check the consistency of the given value based on the number of factor criteria . Consistency should close to zero. Figure 4.14shows the intensity of importance, the derived Factor weight for each factor and Figure 4.15 consistency ratio derived from the tool using the given values.

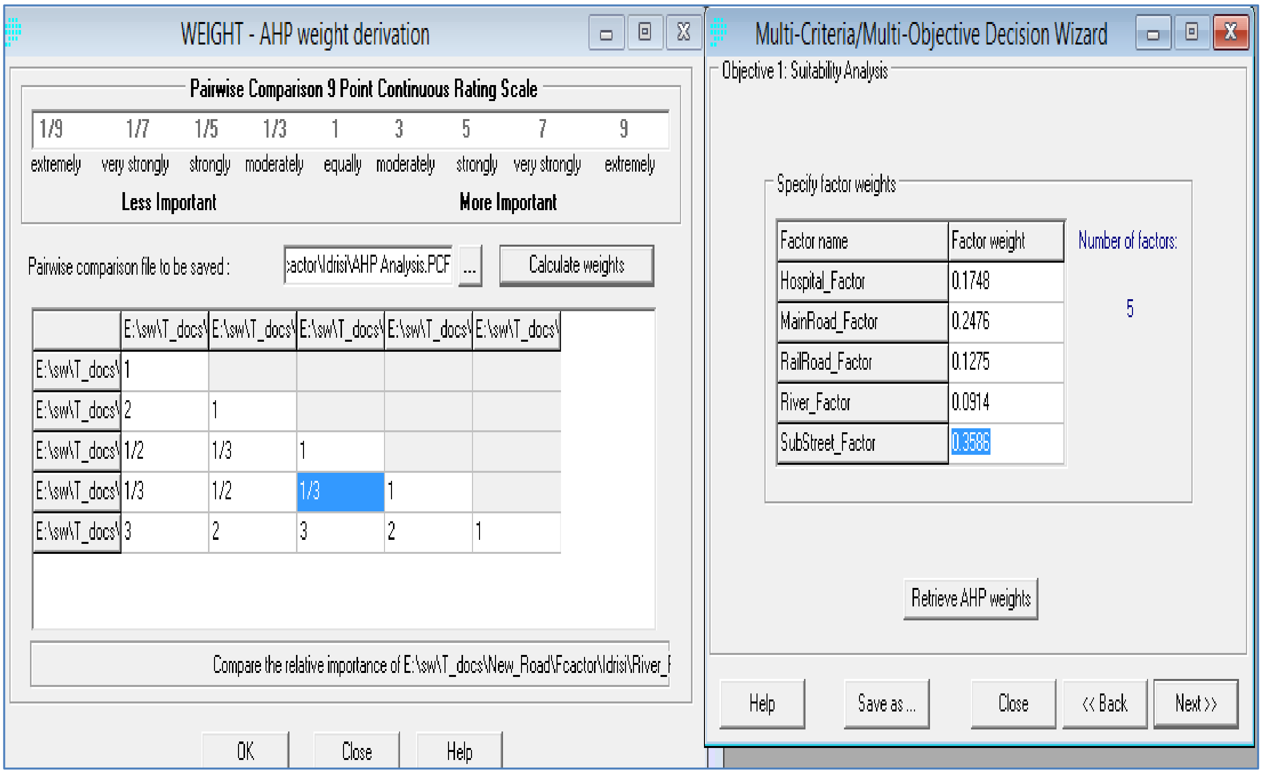

Figure 4.14 : Multi Criteria Decision Wizard

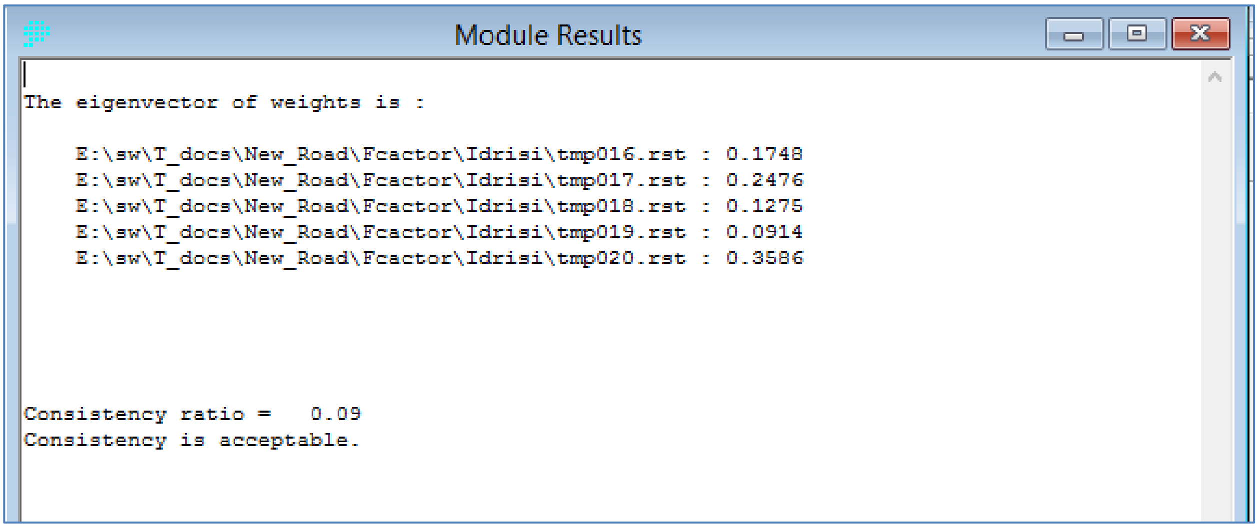

Figure 4.15 : AHP Weight Derivation

4.7 Use of the Discovered Knowledge

After evaluating and finding novel results from spatial data mining for the centrality of the location and the final output of the MCA model for suitability of the site, our model implementation will access and use the cluster centers and overlap with GIS-base MCA result map which used for decision makers to decide based on both outputs. So decision makers, which should select a candidate location before analyzing suitability, will fulfill both the reachability and the suitability of the selected sites at the same time.

CHAPTER FIVE

IMPLEMENTATION AND EVALUATION

This Chapter describes the implementation details of the design of a model for service facility site selection development environment and the tools used to develop the system will be briefly discussed. The implementation details of the group analysis, AHP and implementation of site selection also will be described.

Case Implementation for Hospital Site Selection

As far as we have the necessary standards, procedures, data and defined criteria for any service facility it is possible use the designed model. We selected hospital site selection for the Nifasilik Lafto sub city as a sample for our case implementation because, the Ethiopian standard agency has clearly defined site selection criteria for health facility including hospitals and we have sufficient related works for hospital site selection just for sample using GIS related analysis to adapt the good work.

5.1 Development Tools

The implementation is done using SQL Server 2012 as a spatial database. ArcGIS server to publish both group analysis and suability analysis output found in the database which will later consume by our ArcGIS API. C# programming language using Dojo for spatial computation and displaying spatial layers used for decision making and used as base map .We used Idrisi selva for its interactive decision making wizard. The wizard has complete MCA analysis with AHP for suitability analysis .We used ArcMap for data preparation and pre-processing and group analysis. Our system can run on a single machine which has Intel(R) Core(TM) i7 CPU@ 2.3 GHz each processor, 6.00GB of RAM, 500GB of hard disk, and Microsoft Window 8 operating system.

5.2 The Prototype User Interface

Based on the designed model, prototype system is developed to evaluate the performance and usability of the proposed model. Figure 5.1 shows service facility site selection home page.

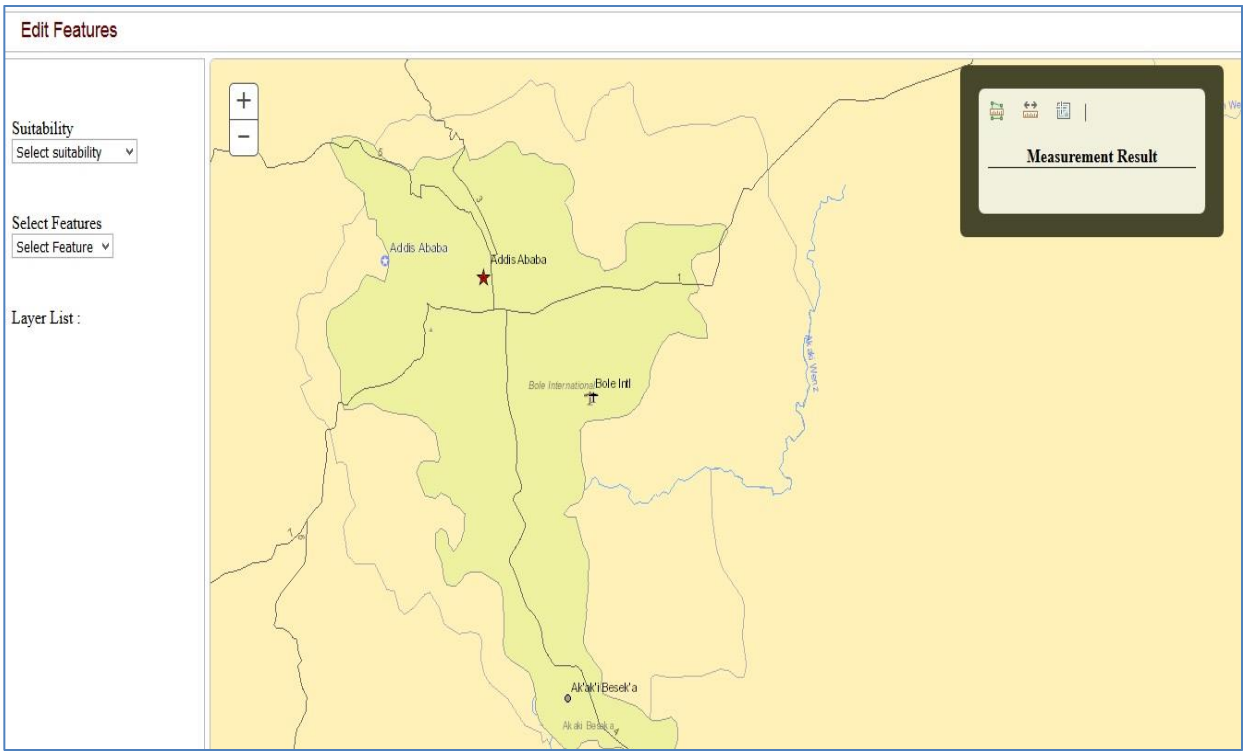

Figure 5.1 : System Interface for Site Selection

To use the prototype system decision makers needs to select suitability dropdown box to access the needed suitability map for the list used as a base for site selection and select features to access the corresponding spatial data mining result with related spatial features. After selecting and overlapping suitability map and data mining results as shown in Figure 5.2 decision makers and domain experts can measure the road distance from cluster center point (influential node of the network) using a measurement tool following the existing road which is adjacent to the suitable location. The shorter the distance to the center point the better area for selection and we can digitize feature layer and fill the necessary attribute information by double clicking the feature or click on an attribute icon on the left side of the selected site using the prototype system and the layer will automatically save to the geo-database for further comparison.

Figure 5.2: Measurement from Cluster Center to Selected Candidate Location

5.3 Suitability and Centrality Maps

The main objective is to select a suitable service site for facilities and, more attention was given to analyze the suitability and centrality or accessibility of the study area. Figure 5.3 shows the final output MCA before classify it into different sub group based on each cell values

Figure 5.3: MCA Final Result

Suitability increase with increase in cell values so we categorized as five different subgroups

Table 5.1: Suitability Classification

Best suitable	Suitable	Good	Fair	Restricted
179-220byte	96-179 byte	41-96 byte	14-41 byte	0-14 byte

In order to examine the nature and spatial extent of the study area, the input maps (Figures 5.4 and 5.5) were classified as suitability map and spatial data mining map. This graphic presentation helps us to understand and enable us to overlap and leverage the two analysis output directly.

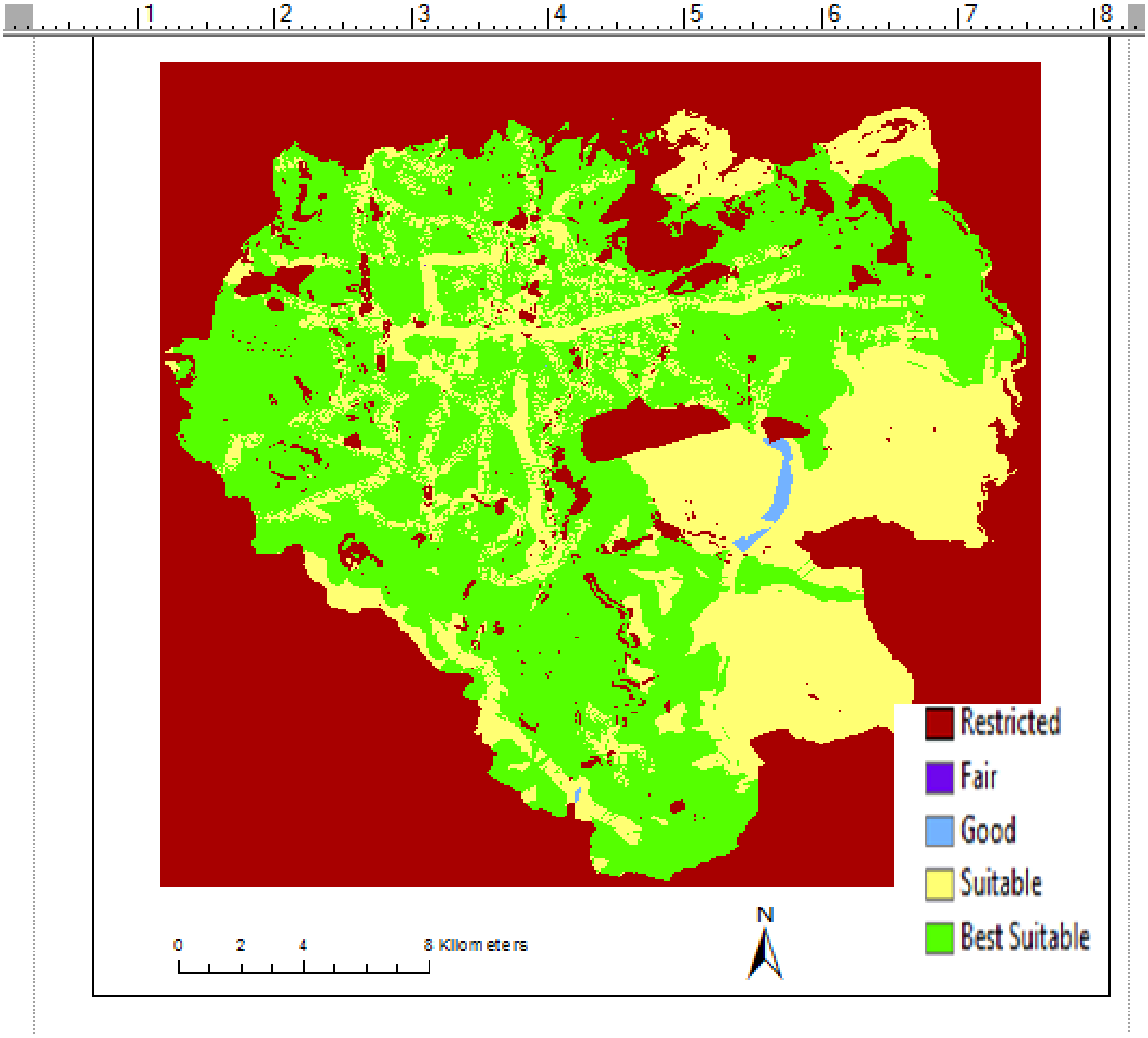

Figure 5.4 : Suitability Map of Addis Ababa

We computed suitability analysis as we described in section 5.3 for the whole Addis Ababa. The need of site selection for the government could vary from the whole area up to kebele level and even service facility type could be any type so, once we computed suitability analysis for the whole area of a single service type and its relative study area could be flexible based on the need of the government or any responsible body. We used Addis Ababa as a whole area, NifasSilikLafto sub city and a single hospital site selection inside it as samples of this work. So as far as we have the boundary and the road data for a specific area, we can conduct our analysis using the overall suitability map. The government may also in need to have more than one service facility in one administrative boundary, in this case, we can group and find the corresponding centers point for each cluster and provide the result for our designed model system which it will handle in any level of administrative area and any kind of service facility given. Figure 6.2 shows center points or influential nodes of Nifasilik Lafto sub city using its own road data inside its boundary.

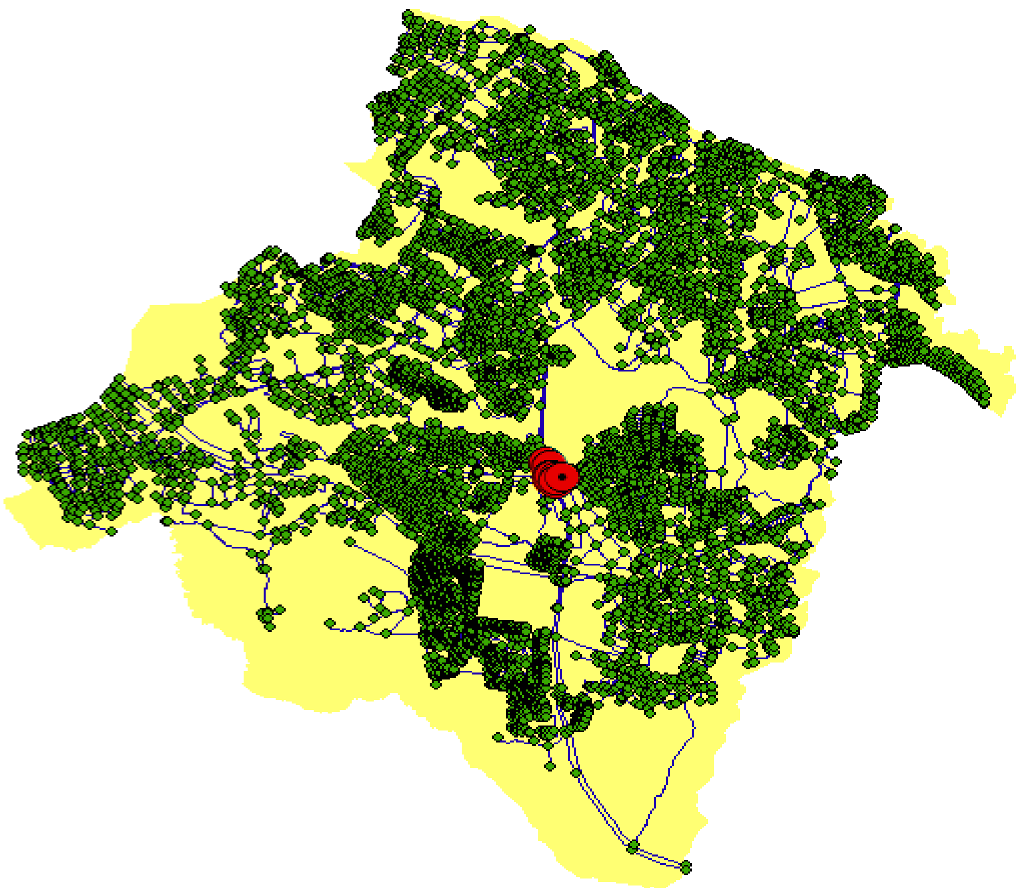

Figure 5.5 : Nifas Silik Lafto Sub City Center Points

5.4 Extraction of Closeness Centrality

In urban network analysis closeness centrality will show or identify which node of the network or spatial points in our case are the best influential or center points from identified groups by computing the distance from each node with the other node. Figure 5.6 shows sample for the computed centrality of each node relative to the other existing node and the final closeness result with its inverse relation as described in **[**29**]**.The smallest result will be the center or the best influential node of the network because the average distance so as we see in Figure 5.7, the minimum value is 0.000028 from the whole computed nodes, we extracted as cluster center by sorting and select relatively smallest values from the computed nodes, which help us to select the best suitable location providing different choice with the same closeness result that makes decision makers more relaxed especially during further analysis of selected candidate location to select specific optimum site.

Table

NifasSilk_Centralty_Featureclass

FID	Shape	ZELEV	UID	SourceID	SourceOID	SnapX	SnapY	Barrier_Co	Closeness
1613	Point	0	1614	2	1614	469236.5474	988169.5215	858993458	0.00002
1614	Point	0	1615	2	1615	469236.8622	989271.0887	858993458	0.000022
1615	Point	0	1616	2	1616	469237.2352	991678.5543	858993458	0.00002
1616	Point	0	1617	2	1617	469237.239	987546.8266	858993458	0.000018
1617	Point	0	1618	2	1618	469238.216	990090.7731	858993458	0.000018
1618	Point	0	1619	2	1619	469238.3672	987738.5	858993458	0.000019
1619	Point	0	1620	2	1620	469239.2447	988787.3025	858993458	0.000022
1620	Point	0	1621	2	1621	469239.8827	990025.3316	858993458	0.000018
1621	Point	0	1622	2	1622	469240.445	991449.9842	858993458	0.000021
1622	Point	0	1623	2	1623	469240.667	987482.6929	858993458	0.000018
1623	Point	0	1624	2	1624	469240.9408	987759.7203	858993458	0.000019
1624	Point	0	1625	2	1625	469241.3079	990729.9423	858993458	0.000022
1625	Point	0	1626	2	1626	469241.9727	987350.5395	858993458	0.000017
1626	Point	0	1627	2	1627	469242.3018	987587.794	858993458	0.000018
1627	Point	0	1628	2	1628	469242.997	990125.7478	858993458	0.000019
1628	Point	0	1629	2	1629	469243.6435	987954.9573	858993458	0.00002
1629	Point	0	1630	2	1630	469243.8126	991883.9527	858993458	0.000019
1630	Point	0	1631	2	1631	469243.94	988592.758	858993458	0.00002
1631	Point	0	1632	2	1632	469244.4559	990540.3064	858993458	0.00002
1632	Point	0	1633	2	1633	469245.0021	988379.3701	858993458	0.00002
1633	Point	0	1634	2	1634	469245.1197	993571.1101	858993458	0.000018
1634	Point	0	1635	2	1635	469245.2861	988795.1483	858993458	0.000022
1635	Point	0	1636	2	1636	469245.7471	991992.4889	858993458	0.000019
1636	Point	0	1637	2	1637	469246.1677	987816.2964	858993458	0.000019
1637	Point	0	1638	2	1638	469246.3632	992321.4622	858993458	0.00002
1638	Point	0	1639	2	1639	469247.0331	991274.6993	858993458	0.000024
1639	Point	0	1640	2	1640	469247.3342	991625.1172	858993458	0.000021

Figure 5.6 : Closeness Centrality Sample Result of Each Node

Figure 5.7 : Best Cluster Center Points from Closeness Centrality Result

5.5 Overlapping suitability and Data Mining Result

Our model for service facility site selection uses both data mining results or cluster centers and suitability result for candidate location selection Figure5.8 shows overlapped suitability map of Addis Ababa as a base map and the cluster center result of NifasSilik lafto sub city for decision makers and domain experts to analyze, measure, digitize candidate locations for further analysis.

Figure 5.8 : Overlapped Suitability and Data Mining Result

5.6 Analyze and Digitize Possible Candidate Locations

Decision makers can select or digitize candidate locations as shown in Figure 5.9 using our model for further analysis.

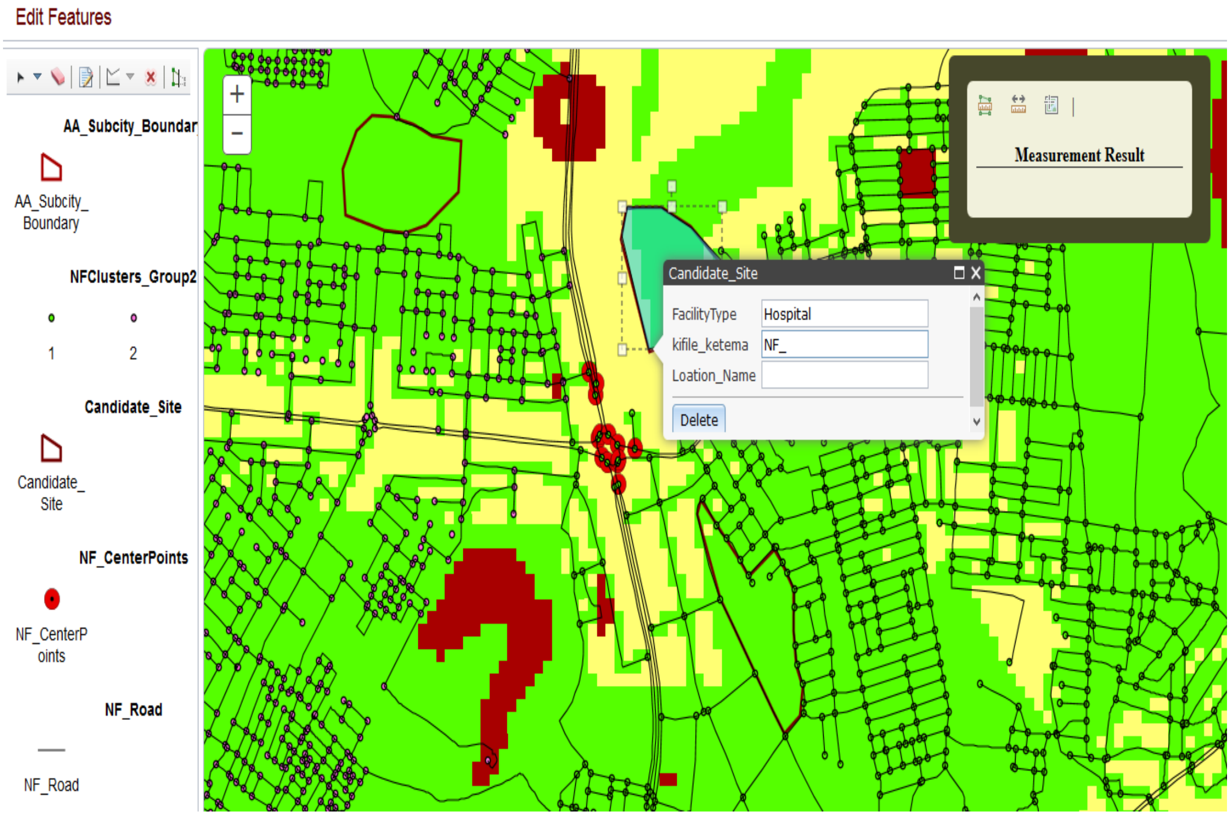

Figure 5.9 : Location Selection and Digitization

Further analysis needed to evaluate and get accurate and reliable result in our model because the two inputs used to select candidate location help experts to identify the location from suitability based on factor criteria and constraints plus centrality or closeness of the location from the surrounding residence using the road network but, because of sensitivity, factor criteria and complexity of AHP weight matrix we can't exhaustively use all factors all together. So after dealing with the main criteria's selected by experts other alternative factor and constraint criteria like land cost, the number of households displaced, distance from main road, distance from residential area and more location specific criteria could come because we can't list all the available Criteria and will be considered during specific site selection by analyzing in-depth the selected candidate sites. Further analysis should be depending on current aerial photo, satellite imagery or using up-to-date full parcel information as a base. In our case we used

Satellite image 2018 of Addis Ababa as shown in Figure 5.10 for further analysis. Our model can consume any kind of spatial information as far as we published it in our registered and known ArcGIS server directory.

Figure 5.10 : Candidate Locations over Satellite Image for Further Analysis

5.7 Further Analysis of Selected Candidate Locations

As we see from the Figure 5.10 with red boundary location one, location two and location three are selected and digitized based on suitability and data mining result all of them have almost equally suitable locations when we see them using Figure 5.9 but when we further analyze them using an up-to-date satellite image they have different suitability as we see the difference in Table 5.2

Table 5.2 **:** Suitability of Selected Location

Location Specific Criteria	**Location one**	**Location Two**	**Location Three**
Distance from Center	600m away	1000m away	300m away
Other Constraints	None	High tension line	None
Other Factors	Relatively close to asphalt	Relatively far away from west discharge	Inside industry area
Displaced Parcels or residence	Only small government shades	None	Both residential and industrial houses will affect
Expansion	Possible	Possible	Needs additional displacement

As we see Table 5.2, we can't select location two because of high tension power line so it should be eliminated. When we see location one it is relatively closer than location two from the center and has small shades which the government can easily shift to another site and it contains relaxed space for expansion to seasonal stream side which we couldn't consider it as a river. When we see the candidate location three it has a big problem like the displacement of many households and government industries which is costly because of compensation and moving the existing industries and residence to another place. So location one will win the bid because of relatively small compensation cost and expandability and probability of making the surrounding area green and clean in the future.

5.8 Digitizing Optimum Site

We adjusted the identified location of the optimum site selection and digitized the best site with 11,280.0 Sq. Meter size as shown in Figure 5.11. The selected parcels are currently using for dairy farm given for small scale industry. We included the whole diary area because for hospital bad odder and west from the farm will affect the whole compound so rather than selecting an adjacent or closer site to the dairy farm it better to change the site for a dairy farm location away with fair compensation or plan for other similar dairy farm location. The final selected site is seen in Figure 5.12 below with green color.

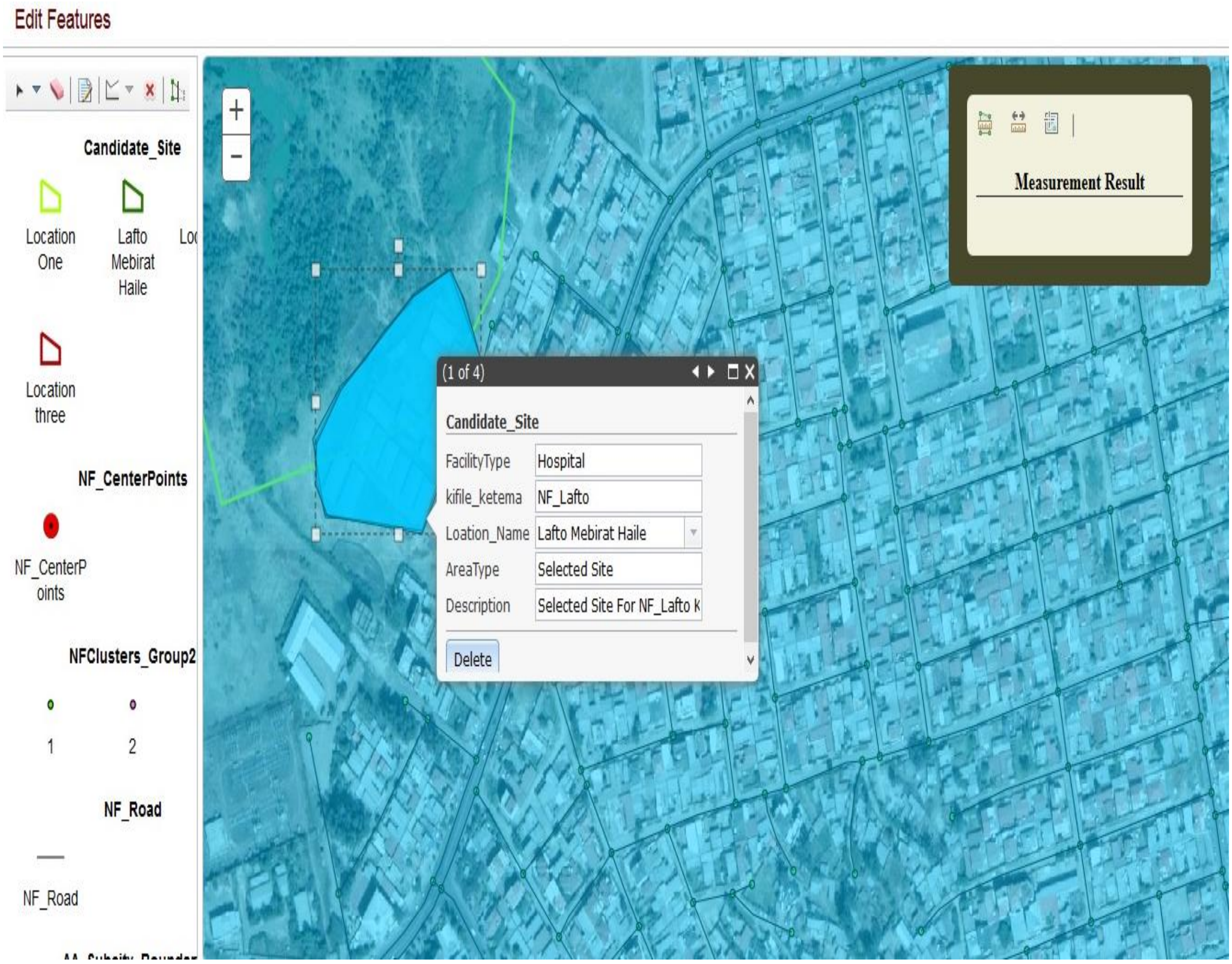

Figure 5.11 : Digitizing Optimum Site

Figure 5.12 : Selected Site Using Satellite Image

5.9 Expert Evaluation

To evaluate the system, we established three groups. The first group contains four IT experts From Information network security agency. The second group contains three GIS experts who have prior site selection knowledge From Information network security agency. The third group contains three urban designers. IT experts and GIS experts already know what is expected from the system. They participated and consulted us during the knowledge engineering phase. But the remaining urban designers did not know what the system is intended to do. Thus, before they started to evaluate the system, we made a detail discussion about what the system is doing. There is a total of ten closed ended questions for each group refer to Annex I. The possible answer for those closed ended questions are, Excellent, Very Good, Good, Fair, and Poor. Therefore, for the comfort of analyzing the relative performance of the system based on users' evaluations, we assigned numbers for each word like Excellent = 5, Very good = 4, Good = 3, Fair = 2 and Poor = 1. The system evaluators gave the value for each closed ended question.

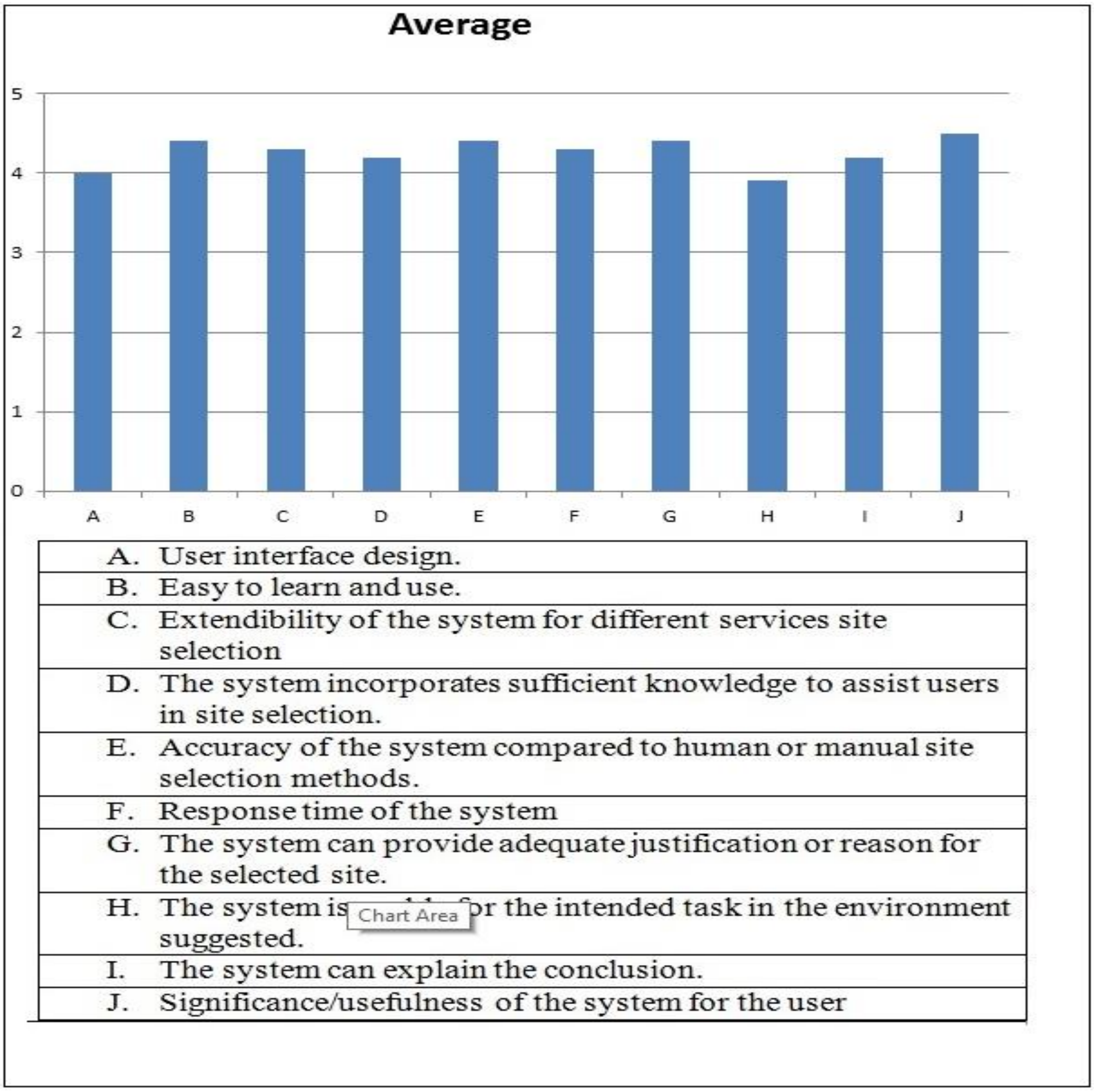

Figure 5.13 : Summary of user's feedback about the proposed model

Depending on the information we gathered from the three validators group, the average percentage of the user satisfaction is about 85.2%. From this average value, we can conclude that the system satisfies the users. And by making some elementary modification, it is possible to increase the satisfaction percentage value. As the scores we collected from the evaluator groups, the effectiveness of the system to provide adequate justification or reason for the selected site is about 88%, refer to Annex II. So, when we evaluate the average percentage value, we believe that we achieve the objective of the research.

5.10 Summary

Service facility site selection prototype system which is developed in this work uses three main inputs as a service from ArcGIS server. Base map, suitability maps and spatial data mining results and contain three main components or sub tasks which are interrelated to each other to extract suitable service facility site. Suitability and features selection component are used to overlap the right suitability map and data mining result features published on ArcGIS sever from dropdown lists. Measurement component is used to measure the exact distance from cluster center to suitable location following the road distance using different measurement units, measuring area of the a polygon and location identification using Latitude and longitude with degree and DMS unit. The edit features component is used to digitize & fill attributes, delete and reshape a selected feature and save updates or digitized feature automatically to geo-database.

CHAPTER SIX

CONCLUSION AND RECOMMENDATION

6.1 Conclusion

In the absence of basic information like spatial distribution about the current existing service facilities site, it would be difficult to determine future planning and improvements for service facility services. This leads to suggest the need to provide up-to date information about land-related resources and different urban network analysis results to help planners in decision-making. Based on the results achieved and analysis done, it is concluded that the objectives of the work were effectively achieved. The following conclusions were deduced from the study. The existence of decision support tools, wizards and different spatial data with the help of GIs desktop software have been provided an opportunity to generate suitability maps. Integrated approaches of suitability analysis map and spatial data mining followed by further analysis with up-to-date image or parcel data are excellent methods to select an optimum site for any service facility. Optimum site selection needs further analysis because of factor and constraints criteria's that couldn't be exhaustively considered because of complexity and different ideas during the selection of factor and constraint criteria. Optimum site selection needs first, candidate location selection with measurable scientific analysis and further analysis of the output using detailed information.

6.2 Contributions of the Work

The main contributions of this research work are given below.

- Identify the major techniques of service facility site selection that should be considered in the development of service facility site selection system.
- Design a model for service facility site selection.
- Identify the critical issues that need to consider for service facility site selection and proposed appropriate techniques
- Select tools that is suitable for optimum service facility site selection

6.3 Recommendations

Up-to-date service facility site selection system needs an up-to-date high quality spatial data. Spatial data collection and pre-processing techniques need qualified professionals like a data collection team in a coordinated fashion with data quality checkers, analysts and domain experts. Hence, we propose the following recommendations for future research direction.

- The developed system in service facility site selection can be improved for better accuracy. If the time and speed in the road network analysis were used or considered spatio-temporal data which, couldn't be achieved without the help of remote sensors for the entire network and careful analysis of road traffic or traffic jam occurs frequently.
- Site selection standards should be measurable in our country. It says "the site should away from noise" which leads into subjective decision and conflicting idea. So, each factor criteria's should be specific and measurable in order our suitability lay on constant and correct criteria rather than using some other country best practice.

[**References**]

[1] City Government of Addis Ababa. ww.aabofed.gov.et. [Online]. www.aabofed.gov.et/Documents/Addis%20Ababa%20Profile%20%202004%20E.pdf, (Accessed on March, 2017).

[2] Krishnamurthy.K,Anjaneyulu. M.V.L.R Sreelekha.M.G, "Interaction between Road Network Connectivity andSpatial Pattern," in *International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST-2005)* , 2016, pp. 131 – 139.

[3] United Nations Human Settlements Programm, "ETHIOPIA:ADDIS ABABA URBAN PROFILE," UN-HABITAT, Nairobi, Kenya, ISBN: 978-92-1-131982-8, 2008.

[4] Joseph P. Messina Pariwate Varnakovida, "Hospital Site Selection Analysis," *Researchgate*, December 2008.

[5] Ilija Cosic, Djordje Lazarevic Aleksandar Rikalovic*, "GIS Based Multi-Criteria Analysis for Industrial Site Selection," in *24th DAAAM International Symposium on Intelligent Manufacturing and Automation, 2013*, Serbia, 2013, pp. 1 0 5 4 – 1 0 6 3.

[6] Jie Wu Lina Zhou, "GIS-Based Multi-Criteria Analysis for Hospital Site Selection in Haidian District of Beijing," högskolan i gävle, Masters Thesis 2012.

[7] Ethiopian Standard Agency. (2012, October) [Online]. http://www.forsslund.org/StandardHealthFaclitiy/General%20Hospita.pdf, (Last visited on March, 2018)

[8] CITY GOVERNMENT OF ADDIS ABABA LAND DEVELOPMENT AND MANAGEMENT BUREAU. (2012, November) Integrated land and Information Center. [Online]. http://www.ilic.gov.et, (Accessed on March 2017).

[9] Busief and Mohamed A. Shouman Ibrahim M, "Hospital Site Selection in Benghazi City in Libya," in *Proceedings of the 2012 International Conference on Industrial Engineering and Operations Management*, Istanbul, 2012.

[10] Mohamed E. El-Sharkawi Mohamed A. El-Zawawy, "CLUSTERING WITH OBSTACLES IN SPATIAL DATABASES," in *Symposium of Signal Processing and Information Technology (ISSPIT01)*, Cairo, Egypt, 2001, pp. 420-425.

[11] Amir Arastoo,Amjad Pirouti Neda Jalaliyoon, "Land Selection; Using Multiple Criteria Decision Making," *International Journal of Academic Research in Management (IJARM)* , vol. 4, no. 1, pp. 14-23, 2015.

[12] Ali A. Alesheikh Saeed Behzadi, "HOSPITAL SITE SELECTION USING A BDI AGENT MODEL," *International Journal of Geography and Geolog*, vol. 2, no. 4, pp. 36-51, 2013.

[13] Magnus Rydén, "Strategic Placing of Field Hospitals Using Spatial Analysis," Royal Institute of Technology (KTH), Stockholm, Sweden, Maters Thesis 2011.

[14] Shawky MANSOUR, "Spatial analysis of public health facilities in Riyadh Governorate, Saudi Arabia: a GIS-based study to assess geographic variations of service provision and accessibility," *Geo-spatial Information Science*, vol. 19, no. 1, pp. 26-38, March 2016.

[15] Man Lung Yiu,Nikos Mamoulis, "Clustering Objects on a Spatial Network," in *international conference on Management of data*, Paris, France, June 2004.

[16] V.K. Bansa Satish Kumar, "A GIS-based methodology for safe site selection of a building in a hilly region," *sciencedirect* , pp. 39–51, January 2016.

[17] Risk and Policy Analysts Ltd (RPA), "Evaluating a multi-criteria analysis (MCA) methodology for application to flood management and coastal defence appraisals," Department for Environment, Food and Rural Affairs , London, Case studies report ISBN 0-85521-146-6, 2005.

[18] Communities and Local Government, "Multi-criteria analysis: a manual," Department for Communities and Local Government, London, Manual ISBN: 978-1-4098-1023-0, 2009.

[19] Fabrizio Battisti, Anthea Chiovitti Maria Rosaria Guarini, "A Methodology for the Selection of Multi-Criteria Decision Analysis Methods in Real Estate and Land Management Processes," *MDPI* , pp. 1-28, February 2018.

[20] Jacek Malczewski Soheil Boroushaki, "Implementing an extension of the analytical hierarchy process using ordered weighted averaging operators with fuzzy quantifiers in ArcGIS," *Computers & Geosciences*, vol. 34, no. 4, pp. 399–410, 2008.

[21] Ali Asghar Alesheikh Farhad Hosseinali, "Weighting Spatial Information in GIS for Copper Mining Exploration," *American Journal of Applied Sciences*, vol. 5, no. 9, pp. 1187-1198, 2008.

[22] Wen-Chin Chen,Ching-Jan Chang Amy H.I. Lee, "A fuzzy AHP and BSC approach for evaluating performance of IT department in the manufacturing industry in Taiwan," *Expert Systems with Applications*, vol. 34, pp. 96–107, 2008.

[23] Benhard Sitohang mam Mukhlash, "Spatial Data Pre-processing for Mining Spatial Association Rule with Conventional Association Mining Algorithms ," in *Proceedings of the International Conference on Electrical Engineering and Informatics Institut Teknologi Bandung, Indonesia June 17-19, 2007*, Indonesia, 2007.

[24] Jamal Shahrabi J Ind Eng Int, "New spatial clustering-based models for optimal urban facility location considering geographical obstacles," *J Ind Eng Int* , March 2014.

[25] Shuliang WANG Deren LI, "CONCEPTS, PRINCIPLES AND APPLICATIONS OF SPATIAL DATA MINING AND KNOWLEDGE DISCOVERY," in *Proceedings of International Symposium on Spatio-temporal Modeling, Spatial Reasoning, Analysis, Data Mining and Data Fusion* , Beijing, 2005.

[26] Michael R. Evans, James M. Kang and Pradeep Mohan hashi Shekhar, "identifying patterns in spatial information: a survey of methods," *Data mining and Knowledge discovery*, vol. 1,

no. 3, pp. 193–214, June 2011.

[27] Pusheng Zhang,Y an Huang,Ranga,Raju V atsavai Shashi Shekhar. citeseerx web site. [Online]. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.77.1454,(Accessed on Aprile, 2017)

[28] D.Volchenkov, "Analysis of urban complex networks," *Condensed Matter Physics 2008*, vol. 11, pp. 331–340, January 2008.

[29] Michael Mekonnen Andres Sevtsuk, "Urban network analysis," in *Proceedings of the 2012 Symposium on Simulation for Architecture and Urban Design*, Florida , 2012, pp. 287-305.

[30] (2017, March) Ethiopian Embassy web Site. [Online]. www.ethiopianembassy.org/AboutEthiopia/InvestmentProjectProfiles/Services/Health/General_Hospital.pdf

[31] Ethiopian Standard Agency. (2018, october) Ethiopian Standard Agency web site. [Online]. http://www.forsslund.org/StandardHealthFaclitiy/General%20Hospita.pdf, (Last visited on March 2018).

[32] Bo Fan, "A hybrid spatial data clustering method for site selection: The data driven approach of GIS mining," *School of International and Public Affairs, Shanghai Jiaotong University, Shanghai 200030*, pp. 3923–3936, 2009.

[33] M. Rahgozar,C. Lucas, F. Taghiyareh A. Zarnani, "patial Data Mining for Optimized Selection of Facility Locations in Field-based Services," in *Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2007)*, 2007, pp. 734-741.

[34] Yonglong Luo,Xintao Ding,Ji Zhang Liping Sun, "A Novel Artificial Immune Algorithm for Spatial Clustering with Obstacle Constraint and Its Applications," *Computational Intelligence and Neuroscience* , vol. 2014, November 2014.

[35] Leisa Armstrong,Barry Croke,Amiya K. T ripathy Sreedhar Nallan, "Geospatial data pre-processing on watershed datasets," in *AFITA 2014*, perth, 2014, pp. 328-336.

Annexes

Annex I: - Questionnaire filled by evaluator groups to validate Service facility site selection model

This questionnaire is used to access the users satisfaction level on the prototype system developed for GIS based urban site selection for hospital services in Addis Ababa city. The result of the questionnaire is purely used for academic purpose and the identity of the responder will not be disclosed.

____

Responder's position or profession: Urban Designers ☐ ; - IT Professional ☐ ;-

GIS Expert ☐

(Put "**x**" mark where it is applicable for you)

Instruction: Please circle the appropriate score corresponding to each question given below. Based on the labeling (Poor = 1; Fair = 2; Good = 3; Very Good = 4; Excellent = 5).

	Evaluation Questions	**Rating Scales**				
1	User interface design.	1	2	3	4	5
2	Easy to learn and use.	1	2	3	4	5
3	Extendibility of the system for different services site selection	1	2	3	4	5
4	The system incorporates sufficient knowledge to assist users in site selection.	1	2	3	4	5
5	Accuracy of the system compared to human or manual site selection methods.	1	2	3	4	5
6	Response time of the system	1	2	3	4	5
7	The system can provide adequate justification or reason for the selected site.	1	2	3	4	5
8	The system is usable for the intended task in the environment suggested.	1	2	3	4	5
9	The system can explain the conclusion.	1	2	3	4	5
10	Significance/usefulness of the system for the user	1	2	3	4	5

Annex II: - The summary of evaluation to validate the proposed model

	Evaluation Criteria	Number of respondents					
		Poor	Fair	good	Very Good	Excellent	Average
1	User interface design.	0	0	3	4	3	4.0
2	Easy to learn and use.	0	0	1	4	5	4.4
3	Extendibility of the system for different services site selection	0	0	2	3	5	4.3
4	The system incorporates sufficient knowledge to assist users in site selection.	0	0	3	2	5	4.2
5	Accuracy of the system compared to human or manual site selection methods.	0	0	1	4	5	4.4
6	Response time of the system	0	0	2	3	5	4.3
7	The system can provide adequate justification or reason for the selected site.	0	0	0	6	4	4.4
8	The system is usable for the intended task in the environment suggested.	0	0	2	7	1	3.9
9	The system can explain the conclusion.	0	0	0	8	2	4.2
10	Significance/usefulness of the system for the user	0	0	1	3	6	4.5
	Overall performance out of 5	**4.26**					
	Overall performance (100%)	**85.2**					

As the scores we collected from the evaluator groups, the effectiveness of the system to provide adequate justification or reason for the selected site is about 88%. This is when we computed the percentage of score in the 7th evaluation criteria.

Declaration

I, the undersigned, declare that this thesis is my original work and has not been presented for a degree in any other university, and that all source of materials used for the thesis have been duly acknowledged.

Declared by:

Name: Firesew Feyiso________________________

Signature: ________________________________

Date: ____________________________________

Confirmed by advisor:

Name: DR. Solomon Atnafu___________________

Signature: ________________________________

Date: ____________________________________