CHAPTER ONE

INTRODUCTION

1.1 Background to the Study

According to Jensen (2007) remote sensing is the science of obtaining information about an object through the analysis of data acquired by a sensor that is not in contact with the remote. The remotely sensed satellite imagery is one of the widely used primary sources of data for land use/land cover classification and change detection analyses. Since aerial photographs came into existence much before multispectral satellite imagery, they have provided past information with high spatial resolution which is invaluable in land use/cover change detection studies.

Increasingly, earth observation has become a prime source of data in the geosciences and many related disciplines enabling research into the distant past, the present and into the future as well. One of the areas of research interest has always been how to relate earth observation output e.g. aerial photographs and satellite images to known features (e.g. land use/cover). In contemporary best practices, the sustainable conservation and preservation of the yields of natural resources are increasingly dependent upon remotely sensed data for inventory natural resources and monitoring of changes (Xie, Sha, and Yu; 2008).

Earth observation is based on the premise that information is available from the electromagnetic energy field arising from the earth's surface or atmosphere (or both) and in particular from the spatial, spectral and temporal variations in that field (Levin, 1999; Kramer, 2002; Sabino, 2005). Through this, the environment can be better monitored and modelled for better developmental policy decisions. A suite of digital data, such as high resolution satellite images is currently available for this purpose. New technologies such as Image Processing (IP), Global Positioning System (GPS) and Geographic Information System (GIS) are being used to integrate and process these data. Digital image processing is extremely important in fully harnessing the information in high resolution satellite imagery data.

Although the aerial photographs usually have relatively high spatial resolution, they lack multispectral information (Tsai, Chou and Wang, 2005). Since the black and white (B&W) aerial photographs came into existence much before the multispectral satellite imagery, they provide past information necessary for land use/cover studies. However, owing to the fact that they lack multispectral information, they cannot be used for automatic land use/cover classification, (Erener and Duzgun, 2009). The leap from manual aerial photographic interpretation to 'automatic' classification was inspired by the availability of experimental data in various bands in the mid 1960's as a prelude to the launch of the Earth Resources Technology Satellite (ERTS-which was later renamed Landsat 1). This necessitated the adoption of digital multivariate statistical methods for the extraction of land cover information (Landgrebe, 1997). Some of the earliest image classifiers at the time included maximum likelihood and minimum distance to means classifiers (Landgrebe, 2005; Wacker and Landgrebe, 1972).

One of the most challenging problems addressed by the remote sensing community in current years is the development of effective data processing techniques for images acquired with the last generation of very high spatial resolution sensors. The development of these kinds of techniques appears even more important in light of recently launched commercial satellites (e.g., Ikonos and Quickbird), with on-board sensors characterized by very high geometrical resolution (from 2.5 to 0.60 m). In this context, great attention is devoted to the analysis of urban scenes, with applications such as road network extraction and road map updating, transportation infrastructure management, the monitoring of growth in urban areas, and discovering building abuse, (Bruzzone and Carlin, 2006).

The leap from manual aerial photographic interpretation to 'automatic' classification was inspired by the availability of experimental data of various band-widths in the mid-1960's as a prelude to the launch of the Earth Resources Technology Satellite (ERTS– which was later renamed Landsat 1) (Maul and Gordon, 1975). Landsat Thematic (TM), a moderate resolution scanner makes it possible to view the co-occurrence of different materials within the ground instantaneous field of view of urban areas that are characterized by a pattern of very heterogeneous patches. This necessitated the adoption of digital multivariate statistical methods for the extraction of land cover information (Landgrebe, 1997). The classification of man-made objects is a basic step in various mapping and modelling applications.

In the recent decades, remote sensing imagery makes the monitoring of the earth surface and atmosphere possible. As the technology of the imagery sensors has improved, the remote sensing images with higher quality have become available. Since extraction of useful information from remote sensing data is important; Scientists manage to propose efficient algorithms for automatic extraction of constructive information from the satellite images. To date, image classification has benefitted from advancements in improved computational power and algorithm development. An example of the subsequent algorithms that have taken root in image classification include k-Nearest Neighbours, Support Vector Machines, Self-Organising-Maps (SOMs), Neural Networks, k-means clustering and object oriented classification. In this regard, support vector machine, originally based on binary function, is viewed as one of the new ways of improving classification accuracies in remote sensing studies (Foody & Mathur, 2004a; Huang, Song, Kim, Townshed, Davis and Masek, 2007). This is because support vector machine (SVM) has the tendency to minimize classification error by minimizing the probability of misclassifying field data drawn randomly from a fixed but unknown probability distribution (Vapnik, 1995, 1998). Support vector machines have been used as a classification method in various domains including and not restricted to species distribution and land cover detection (Rao and Rajinikanth (2014).

To effectively derive reliable information from satellite data, appropriate classification techniques are essential. A number of classification approaches have been developed over the past decades and a review of these algorithms can be found in Lu and Weng (2007). The classifiers can be categorised as either common or advanced. Some of the common classification algorithms include the K-Means, ISODATA, MLC and minimum distance to means (Erdas, 1999; Mather, 2004; Lillesand and Kiefer, 1999; Sabins, 1997; Richards, 1993), while the advanced classification algorithms include the artificial neural networks (ANN), decision trees, support vector machines, and object based image analysis (Lawrence, Bunn, …. 2004; Mahesh and Mather, 2003; Kim, Pang….2003; Mitra and Shankar, 2004; Verbeke, and Vabcoillie, 2004; Foody, 1986; Lucieer, 2008; Hay and Blaschke, 2003; Blaschke and Lang, 2006).

In the work of Anderson, Hardy, Roach and Witmer (1976), a modern nation is like a modern business that requires adequate information on many complex interrelated aspects of its activities in order to make decisions. The knowledge about land use and land cover has become increasingly important as the Nation plans to overcome the problems of haphazard, uncontrolled development, deteriorating environmental quality, loss of prime agricultural lands, destruction of important wetlands, and loss of fish and wildlife habitat. Land use data are needed in the analysis of environmental processes and problems that must be understood if living conditions and standards are to be improved or maintained at current levels. The variety of land use and land cover data needs is exceedingly broad. Land use and land cover data also are needed by Federal, State, and local agencies for water-resource inventory, flood control, water-supply planning, and waste-water treatment. Many federal agencies need current comprehensive inventories of existing activities on public lands combined with the existing and changing uses of adjacent private lands to improve the management of public lands. Federal agencies also need land use data to assess the environmental impact resulting from the development of energy resources, to manage wildlife resources and minimize man-wildlife ecosystem conflicts, to make national summaries of land use patterns and changes for national policy formulation, and to prepare environmental impact statements and assess future impacts on environmental quality.

Consequently, Xie (2006) developed Support Vector Machines as a novel method for land use change modelling with the capacity to effectively address land use change data, which might be a mixture of continuous and categorical variables that are not normally distributed in Calgary, Canada from 1985 to 2001. Radha and Rehna(2014) demonstrated that local invariant features are more effective than standard features such as colour and texture for image retrieval of LULC classes in high-resolution aerial imagery where they performed image retrieval using Multi Support Vector Machine. Bruzzone and Carlin (2006), conducted a Multilevel Context-Based System for Classification of Very High Spatial Resolution Images (panchromatic and pan-sharpened) in order to obtain accurate and reliable maps. Arun and Katiyar(2013) used an Intelligent Approach towards Automatic Shape Modelling and Object Extraction from Satellite Images using Cellular Automata Based Algorithm for accurate detection of features along with automatic interpolation. Ohlhof, Gulch, Muller and Wiedemann(2015), illustrated how Semi-Automatic Extraction of Line and Area Features from Aerial and Satellite Images for 2D and 3D based map were the generation of the military basic vector database using inJECT software extended with new automation modules and a strong GIS interface, Dynamo and Geo-media. Ustuner, Fusun and Barnali (2015), demonstrated the application of Support Vector Machines for Landuse Classification Using High-Resolution RapidEye Images. This research was used in assessing the applicability and sensitivity of SVM in classifying RapidEye images to enhance the agricultural landuse database for the Ministry of Food, Agriculture and Livestock in Turkey.

Okwuashi, McConchie, Nwilo, Isong, Eyo, Nwannekezie, Eyo and Ekpo(2012), predicted the Future Land Use Change of Lagos, Nigeria using Support Vector Machine Based GIS Cellular Automata. Ndehedehe(2014) used Support Vector Machine based kernel types to extract information of the urban areas in Uyo Metropolis from Remote Sensing Multispectral Image using a pixel based classification approach. The United Nations Human Settlements Programme UN-HABITAT (2009) in their Executive Summary of Structure Plans for Awka, Onitsha and Nnewi, adopted the methodology of participatory and very engaging Rapid Urban Sector Profiling for Sustainability (RUSPS) to obtain basic information about the urban services; environment, gender, governance, heritage/historic areas, local economic development; and Shelter and slums of Awka, Onitsha and Nnewi. Igbokwe(2006) while mapping and spatial characterization of major urban centres in parts of South Eastern Nigeria with Nigeriasat-1 Imagery, observed that urban centers in Nigeria are facing the problems of over-stretched infrastructures, environmental degradation, seasonal flooding, destruction of natural vegetation, all resulting from increase in population. To tackle these problems there is need for availability of spatial data, because unfortunately many urban centers in Nigeria lack suitable spatial data. Where they are available, they are grossly outdated. To overcome these handicaps, urban planners are now exploring various new technologies available for quick production of much needed spatial maps.

One of the areas being vigorously pursued is the use of satellite images to generate vital thematic maps of various urban centers. Ojiako and Igbokwe(2009) applied the use of Remote Sensing and Multimedia Geographic Information System (GIS) in the "Administration of Socio-Economic Activities in Nnewi Urban Area of Anambra State". They noted that Nnewi Urban area of Anambra State is faced with a myriad of problems as a result of ineffective management of this spatial information thus leading to loss of enormous revenues by the urban authority,moreover, the available maps are static in form and limited in its application. Eneche, Ahmed and Ekebuike (2018) used the Application of VIS and Land Change Modelers in characterizing and predicting the urban land use/cover dynamics of Nnewi metropolis. Their work was based on a sub-pixel approach classification using Landsat TM/ETM+ satellite imageries of 1986, 2001 and 2016 which were characterized into different LU/LC. However, there is a limited prior research and inconsistency of research resultsin this study area, especially in the LU/LC produced, because it was clarified without considering the whole area of Nnewi-north local government area.

From the above experience, it is obvious that the environment requires constant revisit to be able to capture the temporal changes going on in the environmentusing the earth observation technique which has become a prime source of data in the geosciences and many related disciplines. Nnewi-North Local Government Area in the central region of Anambra State has undergone such rapid urban growth in the last decade. The information obtained from this work will be helpful for suggesting effectiveways of possibly restructuring the development of Nnewi city in a positive direction.

1.2 Statement of the Problem

One of the prime prerequisites for better use of land is information on the existing land use patterns and changes in land cover. In this dynamic situation, accurate, meaningful, current data on landuse/landcover are essential, if public agencies and private organizations are to identify what is happening on the land, and make sound plans for their own future action; then reliable information is imperative. However, this basic geospatial information is relatively lacking in the study area.

Consequently, one important area of application of geospatial technology is in the field of urban and rural mapping, planning and management whether in developed or developing countries. One of the justifications of this research is lack of adequate and correct geospatial information of Nnewi-North Local Government Area, Anambra State, Nigeria. The geospatial information will aid daily planning, development, monitoring and accurate decision-making by end users. Remember, since the history of States and Local government area creations in Nigeria, in 1967, 1976,1987, 1991and 1996 respectively, some of our administrative boundaries were not updated and one of them is the map of old Nnewi Local Government Area.The existing map that are most commonly used for today administrative planning, development, monitoring and decision-making is still showing old Nnewi Local Government Area map comprising of some part of Ekwusigo LGA.To represent the correct land extent of Nnewi-North LGA after Ekwusigo LGA which was carved out of Nnewi-North in 1996 by General Sani Abacha (GCFR) This research will sought and usethe subset of the satellite imagery covering Nnewi-North and South LGA in producing a baseline landuse and landcover map of study area using the correct shape file of Nnewi-North.

Similarly, the development of the city of Nnewi needs serious bottom-up planning rather than top-to-bottom planning approach in order to avoid any avoidable error looming around the city, like unplanned development, narrow roads/poor road network, poor traffic control, lack of open space and gardens among others which obviously are possible because of population growth, urbanization, and industrialization of the zone. From existing information stored in the 1964 topographical map of Nnewi, the city was just a rural area with scattered family settlement pattern, then, population was less and amenities were few or not in existence. But now, the reverse is the case. Therefore, it is never too late to restructure the city for a sustainable urban development.

Nevertheless, Nigerian cities were classified as slums based on western nations' perspective of the development of their cities. (Okpala-Okaka, 2011). Physical observation by this research shows that the development of the city of Nnewi in Southeast, Nigeria seriously needs planning attention in order to avoid acute land use conflicts in the city due to urbanization and industrialization of the area. Land use conflicts in an urban area could endanger human health, safety and lower property values, thus reducing public revenues derivable from property tax system. Before now, Nnewi was just a town or a collection of villages with scattered family settlement pattern, population was less and amenities were few or non-existent, but now the reverse is the case. So, urban renewal and proper planning is required to avoid making the city a chaos of a place. However, planning and development strategies cannot be properly designed or implemented, without first and foremost mapping the existing features within the said area to be able to obtain the topological relationship of features, their shape and size, their accurate position to be able to know the most advantageous ways to forge ahead, bearing in mind the best suitable management option in the development of the area. Therefore, it is vital to restructure the city of Nnewi for sustainable development.

Charles Chukwuma Soludo. (professor of economics and former governor of the Central Bank of Nigeria) in his keynote address at the 2nd Anambra Economic Summit titled "Anambra 2030: Envisioning the African Dubai, Taiwan and Silicon Valley", (Soludo, 2006). Notated that the robuststatisticsonthemajortrendsandfeaturesofAnambraareabsentor weak**.**According to him our major cities have no suitable recreational facility, open spaces, sewage and waste disposal system,not even walkways or flood disposal paths and are not planned for people to live a qualitative life. Rather, it has turned into big, albeit chaotic places, littering of shops and dustbins, leaving our cities with enormous environmental pollutions. In addition, the landmass is so small that the entire state can conveniently be organized into one large free trade zone. He went further to say that Nnewi in particular has lost that momentum of the 1980s and 1990s towards becoming the Taiwan or Japan of Africa. Not even, when Braughtigam had once described Nnewi in an article entitled "industrializing in spite of the state" as the Taiwan of Africa. To him, the failure of Nnewi to sustain the momentum points to the consequences of failure to plan because what you can see today is an array of comatose industries littering Nnewi and its environs which obviously justifies the saying that "if you fail to plan, you plan to fail",(Soludo, 2006).

This research was therefore provoked because of these aforementioned failures, another good example of this urban chaos is the construction of shopping malls within the only available public open space called “the Nkwo Nnewi triangle” which is being used by the people of the area for recreational activities. In view of these, the research seeks a better way to capture the urban features of Nnewi using high resolution satellite imageries and SVM to produce the landuse and landcover map of the area which will serve as a baseline data for urban analysis in Nnewi North Local Government Area of Anambra state, South-Eastern, Nigeria.

1.3. Aim and Objectives

1.3.1 Aim

The aim of this research is to identify and map features (natural and man-made) in high resolution satellite image using Support Vector Machine with a view to developing a reliable urban landuse and landcover map, that will serve as a base map for improved land-use planning and monitoring by end-users.

1.3.2 Objectives

The aim will be actualized through the following objectives:

1. To identify and extract Region of Interests (ROIs) in a subset HRSI of the study area
2. To evaluate the result ofregion of interests and ground truth extraction using the Jeffries-Matusita and Transformed Divergence separability index
3. To perform image classification usingSVM in ENVI 5.1 Software
4. To evaluate the performance of SVM andMaximum Likelihood (ML) in mapping geometric features using Error Matrix, Kappa and a newly proposed Post-Confusion Matrix (PoCoMa) template
5. To produce the landuse and landcover map of Nnewi-North Local Government Area

1.3.3. Research questions

1. What is the best way to identify and extract objects of interest (ROIs), which are both the physical and cultural features of interest in a subset HRSI of the study area using ENVI 5.1 software?
2. What is the performance of the result of region of interests and ground truth extraction using the Jeffries-Matusita and Transformed Divergence separability index?
3. How can you classify man-made and natural objects in high resolution satellite imageries using support vector machine classifiersand Maximum Likelihood Classifier?

4. What is the performance of SVM and MLC in ENVI 5.1 Software window using High Resolution Satellite Imageries (HRSI)?
5. How will the produced landcover and landuse map of Nnewi-North LGA be relevant for a variety of end-users?

1.3.4. Research hypothesis

Hypotheses one

Null hypothesis (H_0): There is no significant difference between the performance of SVM and MLusing Error Matrix, Kappa and a newly proposed Post-Confusion Matrix (PoCoMa) template

1.4 Justification of the Study

It is clear, that "map" as a number one planning tool is rather lacking in this area and when maps and master plan of an area are available, they will definitely guide in urban development, especially as a working document for enforcingplanning laws by both government and the private urban developers. Thus, map is very essential for geospatial location of features and for siting new developments among numerous applications, and it is understandable that sustainable planning and development strategies cannot be properly implemented without mapping the existing features within an area in question to be able to obtain the topological relationships of geospatial features whenever it is required, to be able to obtain factual information for planning, but there is little or no information about all these.

Automatic feature extraction from satellite imagery is cost effective and fast. An essential issue in this context is the degree of accuracy for thematic correctness obtainable through common pixel-based and object-oriented classification algorithms, (Hecher, Filippi, Guneralp and Paulus, 2013). The extraction of objects such as buildings or roads from digital aerial imagery is not only scientifically challenging but also of major practical importance for data acquisition and update of geographic information system (GIS) databases or site models, (Mayer 1999, Baltavias, 2004; Unsalan and Boyer, 2005; Brenner, 2005; Haala and Kada; 2010). Automatic building detection from monocular aerial and satellite images has been an important issue to utilize in many applications such as creation and update of maps and GIS database, change detection, land use analysis and urban monitoring applications. According to rapidly growing urbanization and municipal regions, automatic detection of buildings from remote sensing images is a hot topic and an active field of research, (Ghaffarian and Ghaffarian 2014).

In classification of remote sensing data from urban areas, the identification of relatively small objects, e.g., houses and narrow streets, is important. Therefore, high spatial resolution of the imagery is necessary for accurate classification. Fauvel, Chanussot, and Benediktsson (2009), and Heiden, Segl, Roessner and Kaufmann (2007) analyzed comprehensive urban spectral libraries of anthropogenic and natural materials and point to the value of the high spectral information content for improved separability. Studies by Roessner,SeglHeiden, and Kaufmann. (2001), Herold, Gardner, and Roberts (2003), van der Linden and Hostert (2009) and Franke, Roberts, Halligan and Menz (2009) demonstrated the value of imaging spectrometer data for mapping thematically meaningful urban surface types (e.g., rooftops and paved-areas, bare soil, grass- and tree-cover), or for mapping anthropogenic materials and vegetation species. Extracting man-made objects in satellite images which are generated from the meter to sub-meter resolution plays an important role in remote satellite image analysis. These images contain visual information about various natural and man-made features on or above the surface of the earth.And according to Neelamegam and Ramaraj, (2013) database and information technology has given rise to an approach to store and manipulate thesevital data for further decision making.

Recently, land-use change has been the main concern for worldwide environment change and is being used by city and regional planners to design sustainable cities, Mubea and Menz, 2012). The lack of knowledge relating to land cover and its dynamics especially in developing countries can be attributed to: (1) weak government support for mapping agencies and research institutions, (2) expensive software and hardware, (3) insufficient budget allocations for data purchases and (4) resistance to changes especially by the traditionalist in the field of mapping. However, the increasing availability of inexpensive or free data such as that provided by the global land cover facility (GLCF), the constant drop in the prices of hardware and software as well as improved awareness about the potential applications of remote sensing technology provides the needed momentum for land cover change assessment in the developing world. The combined use of remote sensing and geographical information systems (GIS) will render the essential tools for land cover mapping, storage, analysis and modelling of future scenarios (Geneletti and Gorte, 2003).

In order to portray the significance of this research and products of geospatial information, it is crucial to always consider good planning as avital aspect of urban sustainability, for example, provision of space for urban agriculture, open space, parks/gardens, hotels, modern infrastructure with efficient transport network programs, space for future expansion and longtime city planning and management of the built environment, and this vital information about Nnewi and its environs is very little or totally lacking. In Nigeria, major cities were planned without a strong room for urban sustainability and diversification of economy through harnessing resources domiciled within their immediate environment. Surveying and Geoinformatics is a serious professional task that always use computational and analytical skill to seek after better ways of obtaining reliable information from the geospatial space to provide answers to numerous applications in the world seeking the services of surveyors. This relevant information will be used religiously by urban developers, urban and regional planners, civil engineers, architects, Government and NGOs policy makers, Environmentalists, Estate Surveyors and Valuers, Military, marketers, Transporters, GIS analysts and other stakeholders in the built environment and advocate of sustainable development.

Now, the question that can often arises in developmental planning processes is, how can vital information be extractedfrom land based objects for effective developmental decision making purposes? Extant geospatial practice uses manual digitization which is a time consuming procedure and not appropriate for capturing the changing land details, but in this modern world an analyst needs some fast techniques which can extract the land boundaries, as well as give the information associated with them such as their area and the features along the boundary lines. Thus, proper segmentation or classification of aerial and satellite images of the earth (or geospatial images) are critical sources of information in diverse fields such as geography, cartography, meteorology, surveillance, city planning.

Hence, we can never belabor this point, that the input of experts in surveying and geoinformatics with a view of drawing a road map to aid sustainable planning and urban development isa prerequisite for national development. So, Nnewi the “Japan of Africa” as it is fondly called cannot afford to be left unplanned. The significant role that land use and land cover information plays in environmental monitoring and our knowledge of its dynamics especially in the rural parts of Africa is still lacking. Therefore, landcover information depicting the natural and cultural features that will be mined from high resolution satellite imageries and classify using SVM which will serve as a base map for urban development and possible renewal. Moreover, monitoring the daily development going on in Nnewi city is a necessity.

1.5 Scope of the Research

The study will cover the whole part of Nnewi-North LGA. Machine learning based data classification will be used for identification and mapping of man-made and natural objects in high resolution satellite imageries, using Geometric image features and SVM in ENVI software. The researcher will classify, identify, and map building foot-prints, road networks, streams/rivers and vegetation areas. The automatic interpretation of visual data is a comprehensive task in computer vision field. The machine learning approaches improve the capability of classification in an intelligent way and hence the capability ofSVM in ENVI software will be confirmed using HRSI.

1.6 Limitations of the Study

There are some areas in which the methodology of this research is incapacitated. The created Post Confusion Matrix (PoCoMa) template which was used by the researcher has not really been developed into algorithm. Moreover, the researcher could not access to the right datasets such as high resolution satellite images of various epochs that can cover the researcher's area of interest. This made it impossible to assess the rate of urban dynamism in the study area. Yet again, this affected the production of topographical map of the study area since the surface terrain data could not be extracted from High Resolution Satellite Images used. These could be the shortcomings of the study.

1.7 Study Area

1.7.1 Geographical location

The study area for this research is Nnewi North Local Government Area of Anambra state, South-Eastern zone Nigeria, created in 1996 from old Nnewi Local Government Area. It spans over 1,076 square miles (2,789km^2), lying about 25km south of Onitsha in Anambra state, (Amanze, Ezeh and Okonkwo, 2015). It is located between latitudes 5°59' 41.64"N and 6°03' 28.44"N and longitudes 6°03' 28.44"E and 6°52' 41.64"E (Fig. 1.2a), It is bounded in the north by Idemili South, to South West by Ekwusigo, and South East by Nnewi South L.G. As of Anambra State (Fig. 1.1b). The study area is also known as Nnewi urban and consists of four Quarters (large villages); Otolo, Umudim, Nnewichi and Uruagu (Fig 1.1c). The local government headquarters is located at Umudim; the famous Nkwo market is sited in Uruagu; the Nnamdi Azikiwe University Teaching hospital is situated in Nnewichi, while Otolo ward houses most of the major industries in Nnewi urban area, (Obeta, 2015).

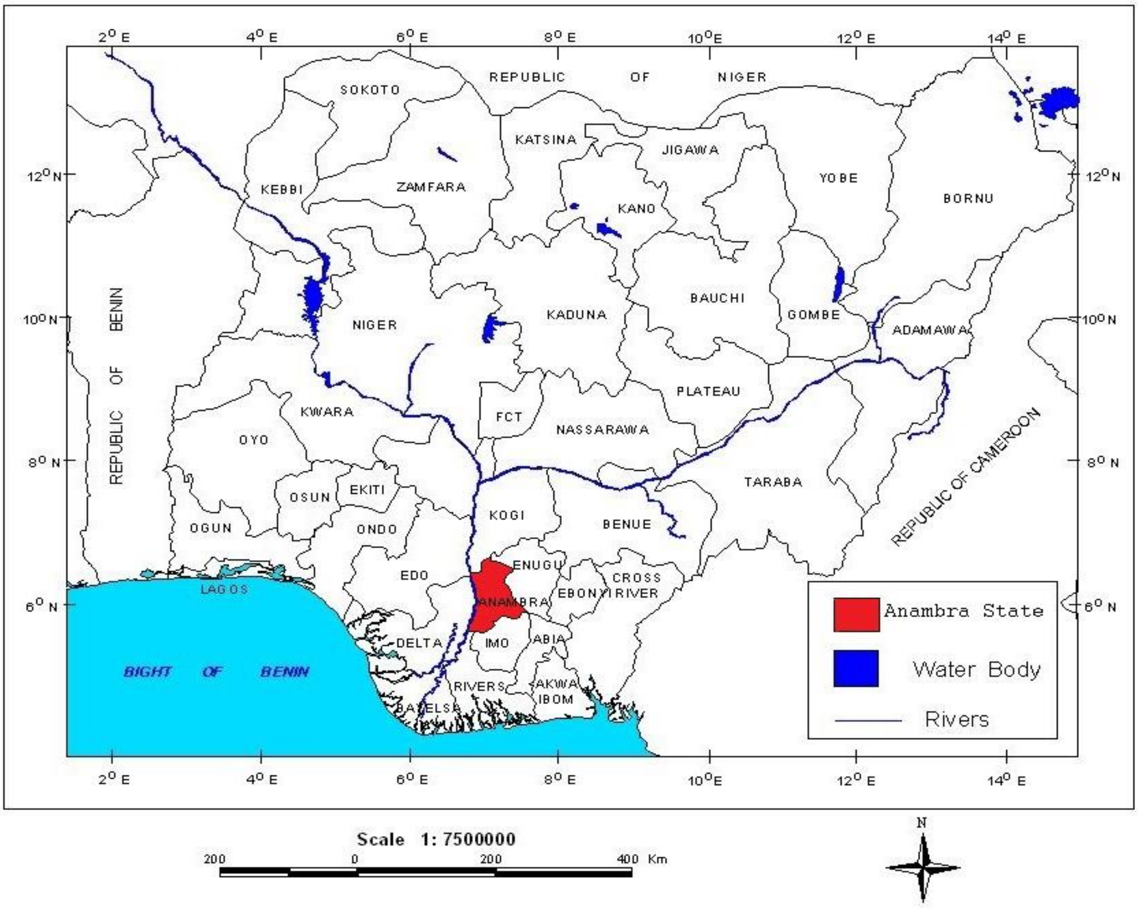


Figure 1.1a: Location of Anambra State in the map of Nigeria

Source: Remote Sensing and GIS Laboratory, Department of Environment Management, Chukwuemeka Odumegwu Ojukwu University, (2016)

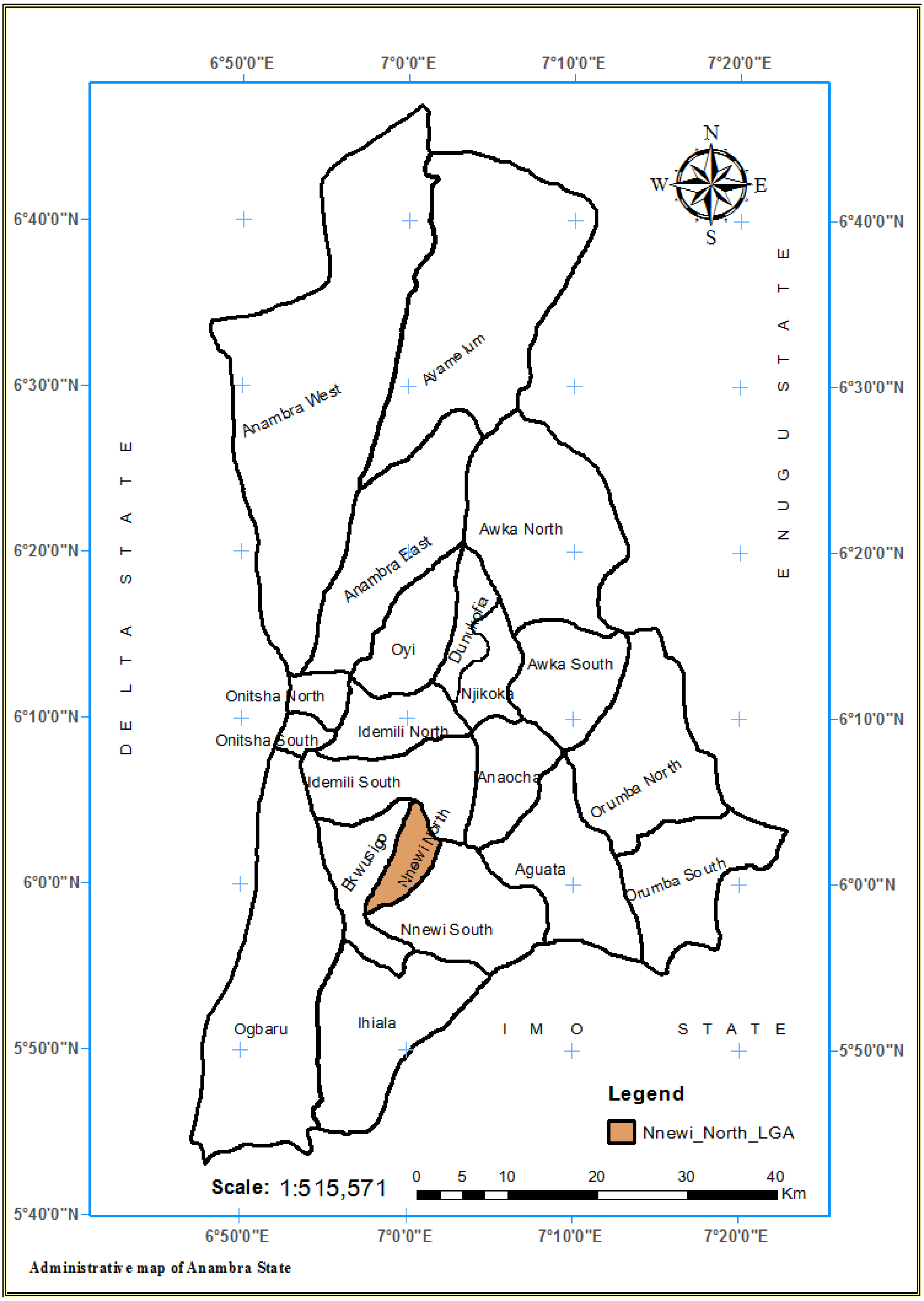


Figure 1.1b: Location of Nnewi North LGAin map of Anambra state

Source: Remote Sensing and GIS Laboratory, Department of Environment Management, Chukwuemeka Odumegwu Ojukwu University, (2016).

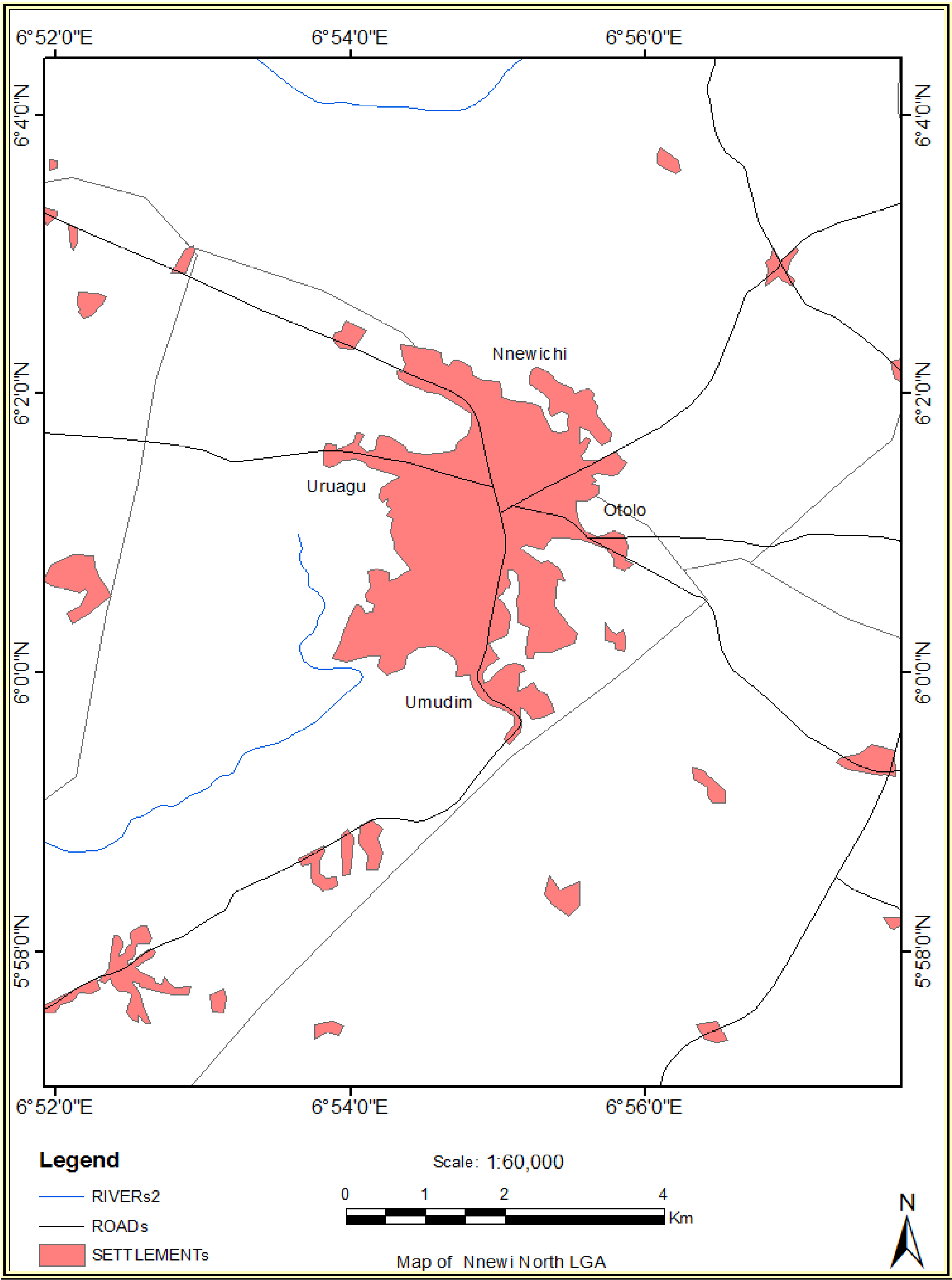


Figure 1.1c: The thematic map of study area depicting the four quarters of Nnewi North

Source: Remote Sensing and GIS Laboratory, Department of Environment Management, Chukwuemeka Odumegwu Ojukwu University, (2016).

1.7.2 Geomorphology of the area

Topographically, Nnewi North is located on the lower Niger plains and Mamu River plains. These plains and lowlands have great rolling topography, generally below 122m above sea level. The Awka-Orlu uplands extend to this area, (Orajiaka, 1975). The altitude ranges from 105 m to 300 m above sea-level. The soil of the area is derived from the sedimentary rocks of middle cretaceous to lower tertiary age and of ferralitic type, (Jungerius, Badwen and Obihara 1964). Although it is rich in free iron, its low mineral reserve and deep porous red colour conveys an illusion of homogeneity of its properties across the area (Phil-Ez, 2010; Obeta, 2015).

1.7.3 Climate and vegetation

The area witnesses 2 distinctive climatic seasons in a year, the dry season and rainy season, with a maximum temperature of 34^{o}C and minimum temperature of 24^{o}C. It has a tropical climate influenced by two major trade winds, the warm moist south west trade winds during rainy season (April-October) and the northeast trade wind during the dry and dust harmattan (November – march), (UN Habitat, 2009). The city records highest rainfall around September. The rain occurs as violent downpour accompanied by thunderstorm and heavy flooding, (Orjiako and Igbokwe, 2009). Humidity of Nnewi North varies between 40% and 92%. It is generally high during the early hours of the day, highest in July during the rainy season and lowest in January during harmattan (Phil-Eze, 2010). The study area falls within the tropical rainforest zone of West Africa, characterized by the presence of many tree species types with the dominant of oil palm trees. This vegetation however, is degraded by over 30 years of infrastructural development, aggressive industrial and urbanization processes, (Orjiako and Igbokwe, 2009, Phil-Eze 2010).

1.7.4 The socio-economic status of the area

The traditional monarch of Nnewi is called the Igwe. The present Igwe is His Royal Highness Igwe Kenneth Orizu III, who presides over the affairs of Nnewi. Otolo, a premiere quarter of the four quarters in Nnewi has been outstanding in all aspects of human endeavors. In it is seated the mantle of leadership that governs Nnewi in consideration with the Obis of the other quarters for the past decades. Its central success is figured in commercial trade but not limited to it, as its cultural heritage has always been the beacon of light to other neighboring villages, (Ogbuagu, 2010). Igwe Kenneth is being assisted by other three Obis of the other three villages. In Nnewi, Ofia-Olu festival is an annual event and all the four quarters observe it. It is part of people’s culture that is very clear to them and affords them an opportunity to relax and enjoy themselves at year’s end. The festivals are usually very colourful and feature a lot of cultural masquerades and dancers (Ogbuagu, 2010). Another festival is Iriji (New yam festival) which last for about a period of the month (August – September).

The population of Nnewi according to NPC (2007) was 79,962 for male and 77,607 for female totalling 157,569. Since the annual growth rate of population in Nigeria is estimated at 3.2%, the population of Nnewi was projected to 2017 as 222.816.The main occupation of the people of the study area is trading and farming, therefore they depend mainly on Agriculture and commerce for their daily livelihood. Most of the prime cash crops produced include; oil palm, raffia palm, groundnut, melon, cotton, cocoa, rubber, maize etc. food crops such as yam, cassava, cocoyam, bread fruit, three-leaf yam etc. Nnewi is home to many agro-allied, automobile and manufacturing industries (Amanze et al., 2015, Anambra RHFA Report, 2013). Landuse within the area is not orderly and varies between residential, commercial, industrial and cultural types with each type dominating according to the perceived interest of the landowners, (Phil-Eze 2010).

Nnewi, popularly called the Japan of Africa is said to be the second largest economic hub of Anambra state after Onitsha and one of the largest in West Africa, (Brautigam 1997, Orji and Obasi, 2012). Together with Aba and Onitsha, Nnewi forms part of the eastern Nigeria’s “New industrial axis”. Nnewi urban is an authentic "manufacturing miracle." Small and medium sized industries have set up in the town and are producing not only for the Nigeria markets but - albeit still to a limited extent - for markets abroad. Industrialization of the town began around 1970 when Nnewi motor parts traders began marketing their own brand name products instead of the reproductions of "original" parts (Onwutalobi, 2009). Some are confident enough to call it “Nigerian Japan” (Brautigam 1997).

The inhabitants are predominantly traders and manufacturers of auto and auto spare parts. It plays a leading role as a centre for the assembly and distribution of motorbikes, spare-parts and other business activities in Nigeria. it has institutions like: Nnamdi Azikiwe University Teaching hospital, Nnewi; College of Health Sciences Nnewi, Odumegwu Ojukwu Polytechnic, Ezinifite, Nnewi South among others, (Anigbogu, Onwuteaka, Edoko and Okoli, 2014).

CHAPTER TWO

LITREATURE REVIEW

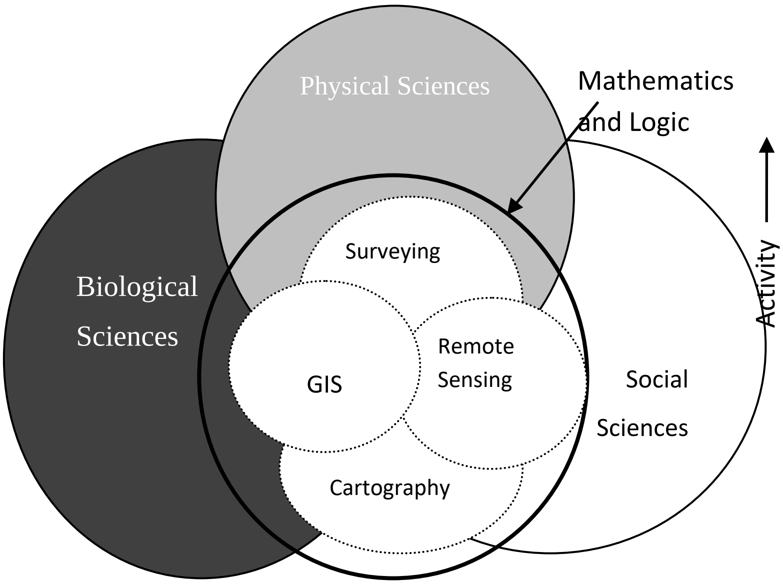
This chapter presents the principles, theoretical and conceptual frameworks of Remote Sensing, feature extraction, spatial analysis and Support Vector Machine, including related theorieson data mining, image classification and image resolution used in this study. Nevertheless, the literature of various scholars relating to the concepts of this study is also presented.

2.1Theoretical Framework

2.1.1 Remote sensing and the interaction model

Remote sensing is a tool or technique similar to mathematics. Using sophisticated sensor to measure the amount of electromagnetic energy exiting an object or geographic area from a distance and then extracting valuable information from the data using mathematically and statistically based algorithms is a scientific activity (Fussell, Rundquist and Harrington, 1986). A science is defined as broad field of human knowledge concerned with facts held together by principles. Interestingly, some persons do not consider mathematics and logic to be science, but the fields of knowledge associated with mathematics and logic are such valuable tools for science that we cannot ignore them. Remote sensing functions in harmony with other geographic information sciences often referred to as GIScience, including cartography, surveying and geographic information systems (GIS) (Curran, 1987; Clarke, 2001; Jensen, 2005).

Dahlberg and Jensen (1986) and Fisher and Lindenberg (1989) suggested a model where there is interaction between remote sensing, surveying, cartography and GIS, where no sub-discipline dominates and all are recognized as having unique yet overlapping areas of knowledge and intellectual activity as they are used in physical, biological and social science research. See interaction model (fig. 2.1) below depicting the relationship of the geographic information sciences (remote sensing, geographic information systems, cartography and surveying) as they relate to mathematics and logic and the physical, biological and social sciences.



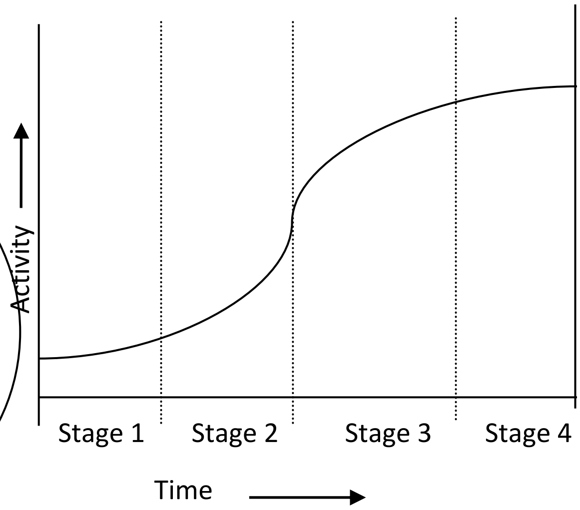


Figure 2.1: Interaction Model (Source: Jensen 2007)   Figure 2.2: Developmental Stage (Source: Jensen 2007)

The theory of science suggests that scientific discipline go through four classic developmental stages. Wolter (1975) suggested that the growth of a scientific discipline, such as remote sensing that has its own techniques, methodologies and intellectual orientation seems to follow the sigmoid or logistic curve illustrated in figure 2.2. The growth stages of a scientific field are: stage 1- a preliminary growth stage of with small increment of literature; stage 2 – a period of exponential growth when the number of publications doubles at regular intervals; stage 3 – a period when the rate of growth begins to decline but annual increments remain constant; and stage 4 – a final period when the rate of growth approaches zero.

Using this logic, it may be suggested that remote sensing is in stage 2 of a scientific field, experiencing exponential growth since the mid-1960s with the number of publications doubling at regular intervals (Colwell, 1983; Cracknell and Hayes 1993; Jensen 2005). We may be approaching stage 3 with increasing specialization and theoretical controversy. However, the rate of growth of remote sensing has not begun to decline. In fact, there has been a tremendous surge on the number of persons using remote sensing and commercial firms using remote sensing during the 1990s and early 2000s (Davis, 1999; ASPRS, 2004). Significant improvements in the spatial resolution of satellite remote sensing (e.g. more useful 1 x 1 m panchromatic data) has brought even more social science GIS practitioners into the fold. Hundreds of new peer-reviewed remote sensing research articles are published every month.

Remote sensing (RS), also called earth observation, is the acquisition and analysis of information about the earth from a distance using a computer and sensor through electromagnetic radiation. This started in 1830s with the origination of the camera (Jorgensen, 2004). Earth observation is based on the premise that information is available from the electromagnetic energy field arising from the earth's surface or atmosphere (or both) and in particular from the spatial, spectral and temporal variations in that field (Levin, 1999; Kramer, 2002; Sabino, 2005). Through this, the environment can be better monitored, modelled, and consequently, better policy decisions made. The Remote Sensing is basically a multi-disciplinary science which includes a combination of various disciplines such as optics, spectroscopy, photography, computer, electronics and telecommunication, satellite launching etc. All these technologies are integrated to act as one complete system in itself, known as Remote Sensing System.

Remote sensing techniques allow taking images of the earth surface in various wavelength region of the electromagnetic spectrum (EMS). One of the major characteristics of a remotely sensed image is the wavelength region it represents in the EMS. Some of the images represent reflected solar radiation in the visible and the near infrared regions of the electromagnetic spectrum, others are the measurements of the energy emitted by the earth surface itself i.e. in the thermal infrared wavelength region. The energy measured in the microwave region is the measure of relative return from the earth's surface, where the energy is transmitted from the vehicle itself. This is known as active remote sensing, since the energy source is provided by the remote sensing platform. Whereas the systems where the remote sensing measurements depend upon the external energy source, such as sun are referred to as passive remote sensing systems. (See Fig. 2.3 for concept of Remote Sensing)

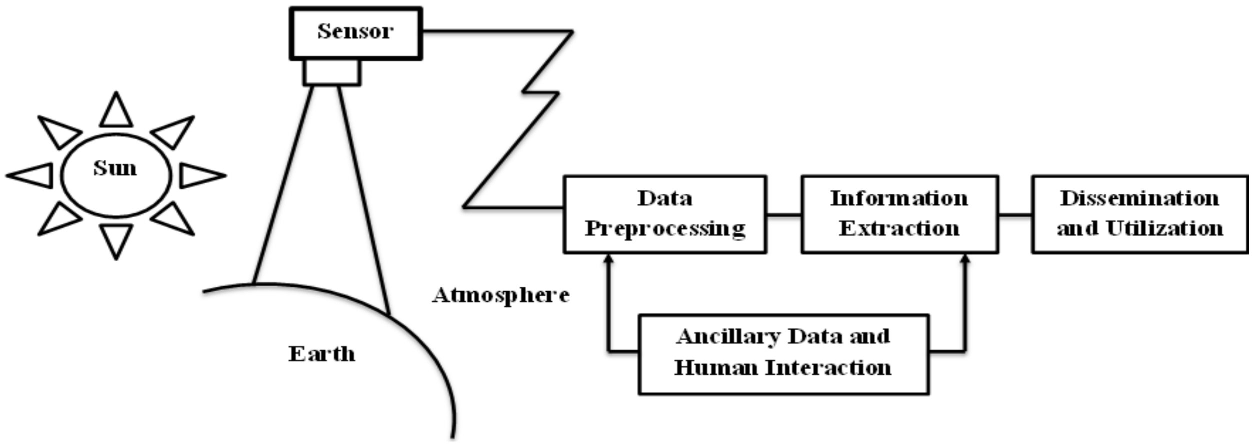


Figure 2.3: Concept of remote sensing (Landgrebe, 1998).

2.1.1.1*Image resolutions*

Spectral resolution is the ability of a sensor to produce clear or distinguished wavelength interval, known as channels or bands in the electromagnetic spectrum. The arrangement of pixels in an image describes the spatial structure while radiometric describes the actual information contained in the image. The temporal resolution is the length of time taken by the satellite to complete one entire orbit cycle. The image obtained by remote sensors may contain one spectral band called panchromatic image (black and white), multispectral images contains few spectral bands and contiguous spectral bands are hyperspectral images. Different class labels and details in an image can be distinguished when their responses over a distinct wavelength range are compared.

2.1.1.2 *Very high resolution satellites imaging systems (VHR)*

VHR images became available (and popular) with the Launch of commercial satellite like IKONOS and Quickbird, with on-board multispectral scanners characterized by a geometrical resolution in the order of 1m. These satellites can acquire four multispectral bands, in the visible and near infrared spectral ranges, and a panchromatic channel with four times higher spatial resolution. These satellites represent a significant improvement in the geometric resolution with respect to the popular Landsat satellites. Indeed, Landsat 7 (the last satellite of the Landsat program) provides seven multispectral bands in the visible, near and thermal infrared ranges with a geometric resolution of 30 m (except the thermal infrared band that has a resolution of 60 m) and a panchromatic channel with a spatial resolution of 15 m. The SPOT 5 satellite, the last launched and operating satellite of the SPOT program, can acquire four multispectral bands in the ranges of visible, near and mid-infrared with a spatial resolution of 10 m (except the mid infrared band that has a resolution of 20 m) and a panchromatic band with a maximum of 2.5 m (Persello, 2010).

Recently, a new generation of VHR satellite systems became available, i.e., GeoEye-1, World-View-1 and 2, which further improve the geometric resolution, providing a panchromatic channel with a resolution of smaller than half meter. It is interesting to note that the World-View- 2, satellite increase the spectral resolution other than the geometric resolution, by providing eight channels instead of the common four. Moreover, in the next years the quality and the availability of this type of data are going to further increase thanks to the missions GeoEye-2 and Pleiades. Table 2. 1 shows the main characteristics of the most popular satellite systems of the last decade with on board multispectral scanners. Fig. 6 shows a graph of the increase of the spatial resolution of popular satellites with on board multispectral systems since 1970.

Table 2.1: Main characteristics of multispectral sensors on board of satellite platform. All the considered satellites are in a polar sun-synchronous orbit with equatorial crossing time 10a.m.

Satellite	Sensor Bands (nm)	Spatial resolution (at nadir)	Swath width	Orbit altitude	Year of Launch
Landsat 7	520-900 (pan) 450-520 (blue) 520-600 (green) 630-690 (red) 760-900 (NIR) 1550-1750 (MIR 1) 10400-12500 (TIR) 2080-2350 (MIR 2)	15 m 30 m 30 m 30 m 30 m 30 m 60 m 30 m	185 km	705 km	1999
Ikonos	526-900 (pan) 445-516 (blue) 505-595 (green) 0.632-0.698 (red) 0.757-0.853 (NIR)	0.82 m 3.2 m	11km	681 km	1999
Eros A	500-900 (pan)	1.8 m	14km	480km	2000
Quickbird	445-900 (pan) 450-520 (blue) 520-600 (green) 630-690 (red) 760-900 (NIR)	0.61 m 2.44 m	16.5 km	450 km	2001
	480-710 (pan)	2.5 m			

Table 2.1. Continued…

Spot 5	500-590 (green) 610-680 (red) 780-890 (NIR) 1580-1750 (MIR)	10 m 10 m 10 m 20 m	60 km	832 km	2002
Eros B	500-900 (pan)	0.7 m	7 km	600 km	2006
World-View 1	450-900 (pan)	0.50 m	17.6 km	496 km	2007
GeoEye 1	450-900 (pan) 450-520 (blue) 520-600 (green) 625-695 (red) 760-900 (NIR)	0.41 m 1.65 m	15.2 km	681 km	2008
Satellite	Sensor Bands (nm)	Spatial resolution (at nadir)	Swath width	Orbit altitude	Year of Launch
WorldView 2	450-800 (pan) 400-450 (coastal) 450-510 (blue) 510-580 (green) 585-625 (yellow) 630-690 (red) 705-745 (red edge) 770-895 (NIR 1) 860-1040 (NIR 2)	0.46 m	16.4 km	770 km	2009
Pleiades-HR 1	480-830 (pan) 430-550 (blue) 490-610 (green) 600-720 (red) 750-950 (NIR)	0.7 m 2.8 m	20 km	694 km	2010
Pleiades-HR 2	480-830 (pan) 430-550 (blue) 490-610 (green) 600-720 (red) 750-950 (NIR)	0.7 m 2.8 m	20 km	694 km	2010
GeoEye 2	Pan	0.25 m			2012

Source: Adopted from(Persello, 2010).

VHR images allow one of the precise recognition of the shape and the geometry of the objects present on the ground as well as the identification of the different land-cover classes. For these reasons, VHR data are very important sources of information for the development of many applications related to the natural environment and human structures (Persello, 2010). Today there are many satellite sensors which provide high resolution images such as IKONOS, Quickbird, GeoEye, and World-View, among others. An elaborate comparison to the state-of-the-art image classification algorithms is recently going on in field of geoscience.

2.1.1.3 *Hyperspectral imagery systems*

Recent developments in sensor technology have resulted in the development of hyperspectral instruments. The instruments are capable of collecting hundreds of images (spectral bands) corresponding to wavelength channels, for the same area of the earth's surface (Green, Eastwood, Sarture, Chrien, Aronsson, Chippendale, Faust, Pavri, Chovit, Solis, Olah and Williams, 1998; Plaza, Martinez, Plaza and Perez, 2003). Each pixel contains in hyperspectral data cube is linked to spectral signature or fingerprint that uniquely characterize the materials within the pixel. Such recognition provides a great advantage for detecting minerals, urban planning and vegetation studies, monitoring and management of environment, security and defense matters among others (Varshney and Arora, 2004). However, accurate classification of remote sensing images is an important task to be able to achieve these advantages (Shaw and Burke, 2003).

In hyperspectral imagery, a pixel is usually mixed with a number of materials present in the scene. Spectral mixture analysis has been extensively used in remote sensing for material discrimination, classification and detection. Various spectral mixture techniques have been reported in the remote sensing literature (Plaza, Martinez, Perez and Plaza, 2002; Keshava and Mustard, 2002; Plaza, Martinez, Perez and Plaza,2004b; Pinho, Silva, Fonseca and Monteiro, 2008; Zhang and Huang, 2010). However, the spectral signature of a particular pixel contains a mixture of the signatures (fingerprints) of the numerous materials present within the spatial coverage of the target field view by the sensor. Spectral unmixing is the process whereby the measured spectrum of a mixed pixel is broken down into a number of pure spectral components, referred to as endmembers. This is also known as class labels, class types, components or signatures (Gong and Zhang 1999) and a set of corresponding fractional abundance that indicate the amount of each endmember present in the pixel (Plaza et al., 2004b; Sanchez, Martin, Plaza and Chang, 2010). Linear spectral unmixing is a commonly accepted approach to mixed-pixel classification. Distinct substances such as water, tree, bridge, grass among others which are called the endmembers and the fraction in which they emerge in the mixed pixel is referred to as fractional abundance. The other method used is the nonlinear mixing, where the incident sun ray comes across close mixture that causes multiple bounces.

2.1.1.4 *Image processing and classification*

Image processing is any form of signal processing for which the input is an image, such as satellite images, a photograph or video frame. The output of image processing may be either image, a set of characteristics or parameters related to the image. Image processing is of interest because it affords abundant data to be translated into useful information in time (Liu, Fan Deng and Ji, 2009). Numerous sources of imagery are identified by their differences in spectral, spatial, radioactive and temporal characteristics, hence are suitable for different purposes of vegetation mapping. Then, correlations of the vegetation types (communities or species) within the classification system with discernible spectral characteristics of remote sensed imagery have to be identified. These spectral classes of the imagery are eventually translated into the vegetation types in the image interpretation process, which is also called image processing (Xie et al., 2008).

Image classification is the process whereby pixels in an image are automatically categorized into land cover classes. It is a fundamental analysis technique for remotely sensed data and involves the categorization of pixels based on their spectral characteristics (Cihlar, Xiao, Chen, Beaubien, Fung and Latifovic, 1998). Image classification can be categorized into (i) supervised and unsupervised, (ii) Spectral and contextual classifications. Supervised classification requires the user to decide which classes exist in the image, and then delineate samples of these classes. These samples (known as training areas) are input into a classification program, which produces a classified image. The choice of the training area is based on the researcher's familiarity with the geographical area and knowledge of the actual land cover types in the image. Unsupervised classification does not require training areas. Actually, it is the opposite of the supervised classification process. Spectral classes are grouped based significantly on the numerical information in the data and then matched by the researcher to the information classes (Bortolot, 1999; Sabino, 2005). Types of supervised classifiers include Minimum – Distance to Means, Neural Networks, and maximum likelihood classifiers, while examples of unsupervised classifiers include K – Mean, Fuzzy C means, and ISODATA among others (Bortolot, 1999; Sabino, 2005).

Spectral techniques are based on the spectral response pattern of a pixel and are divided into two categories, parametric and non-parametric classifiers. Parametric classifiers assume a

Gaussian distribution of the data. In supervised parametric classification, a multivariate Gaussian distribution associated with each class is extracted from the training set by estimating the mean vector and the covariance matrix. Parametric classifiers are based on the definition of some discriminate functions based on the parameters of normal distribution. An example of parametric classifiers is a Gaussian maximum likelihood classifier. Non-parametric classifiers imply decision boundaries of arbitrary geometry and needs an iterative process to complete the boundaries. Example of these classifiers includes Nearest Neighbor (1-NN) 3, K-Nearest Neighbor (k-NN) 4 techniques, kernel methods, and classification trees. Contextual classifiers consider the special context of a pixel in the image and are generally applied on remote sensing data when a large variety of spectral responses are observed in the same field. Contextual classifiers have been used successfully in a number of different problems such as coping with segmentation and classification of remotely sensed data (Dwivedi, Kandrika and Ramana, 2004).

2.1.2 Feature extraction model

Road network, building footprints and vegetation areas are also other important layers in GIS. The extraction of these layers was mostly done by manual digitization. It was not only time consuming but also an expensive process to obtain all the layers. Therefore, automated and semi-automated extraction of urban geospatial information systems has been developed to solve these problems. However, it is still an important and open area to find the automated methods that are accurate and easy to implement properties. Nowadays, the availability of free and commercial high-resolution remote sensing multispectral imagery from sensors such as IKONOS increases the significance of this concept. Theoretically, the detailed depiction of objects and ground surfaces by the dense point cloud may indicate a directness or simplicity in which objects and related information can be retrieved from the data, (Zhang, Lin and Ning, 2013).

Guyon and Elisseeff, (2003) described feature extraction as the transformation of original features to construct a new feature space. Examples of linear feature extraction models are the Principal Component Analysis (PCA) (Hyvarynën, 2001), Independent Component Analysis (ICA) (Fukunaga, 1990) and Linear Discriminant Analysis (LDA) (Hyvarynën et al., 2001) among others. These algorithms minimize some criterion function like the mean square error (PCA), a class separability criterion (LDA) or an independence criterion (ICA). Various classification models use the linear features to build a classifier, obtaining improved computational efficiency and accuracy (Grana and d'Anjou, 2004). Another approach to feature extraction is the linear spectral mixing analysis. A standard technique for spectral mixture analysis is linear spectral unmixing, (Hein and Chang, 2001; Plaza et al., 2004b), which assumes that the collected spectra at the spectrometer can be expressed in the form of a linear combination of end members weighted by their corresponding abundances.

2.1.3 Data mining

Data mining is a logical process that is used to search through large amount of data in order to find useful data (Otukei and Blaschke, 2010). The goal of this technique is to find patterns that were previously unknown. Once these Patterns are found they can further be used to make certain decisions for development. Three steps involved are exploration, pattern identification and deployment. In the first step of data exploration, data is cleaned and transformed into another form, and important variables and then nature of data based on the problem are determined. Once data is explored, refined and defined for the specific variables the second step is to form pattern identification. Identify and choose the patterns which make the best prediction. Patterns are deployed for desired outcome. Data mining algorithms can follow three different learning approaches: supervised, unsupervised, or semi-supervised. In supervised learning, the algorithm works with a set of examples whose labels are known. The labels can be nominal values in the case of the classification task, or numerical values in the case of the regression task. In unsupervised learning, in contrast, the labels of the examples in the dataset are unknown, and the algorithm typically aims at grouping examples according to the similarity of their attribute values, characterizing a clustering task. Finally, semi-supervised learning is usually used when a small subset of labelled examples is available, together with a large number of unlabelled examples.

Data mining concept is growing fast in popularity, it is a technology that involving methods at the intersection of (Artificial intelligent, Machine learning, Statistics and database system), the main goal of data mining process is to extract information from a large data into form which could be understandable for further use. Some algorithms of data mining are used to give solutions to classification problems in database. An algorithm is a computational procedure which takes some value or set of value as input and generates some value or set as output. The result of a given problem is the output that they got after solving the problem. An algorithm is considered to be correct, if for every input for instance, it generates the correct output and it gets terminated and give the desired output otherwise it does not consider as a correct algorithm.

2.1.4 Some common algorithms used in data mining

2.1.4.1 *Decision tree*

A Decision Tree Classifier consists of a decision tree generated on the basis of instances. A decision tree is a classifier expressed as a recursive partition of the instance space. In a decision tree, each internal node splits the instance space into two or more sub-spaces a certain discrete function of the input attributes values. The root and the internal nodes are associated with attributes; leaf nodes are associated with classes. Basically, each non-leaf node has an outgoing branch for each possible value of the attribute associated with the node. To determine the class for a new instance using a decision tree, beginning with the root, successive internal nodes are visited until a leaf node is reached. At the root node and at each internal node, a test is applied. The outcome of the test determines the branch traversed, and the next node visited. The class for the instance is the class of the final leaf node. The process of building the decision tree is presented in Quinlan (1993). The performance of DTs can be affected by a number of factors including: pruning and boosting methods used and decision thresholds (Mahesh and Mather, 2003).

2.1.4.2. *K-nearest neighbour classifiers (KNN)*

The k-nearest neighbour' algorithm is amongst the simplest of all machine learning algorithms. An object is classified by a majority vote of its neighbour, with the object being assigned to the class most common amongst its k nearest neighbour. K is a positive integer, typically small. If k = 1, then the object is simply assigned to the class of its nearest neighbour. In binary (two class) classification problems, it is helpful to choose k to be an odd number as this avoids tied votes. The same method can be used for regression, by simply assigning the property value for the object to be the average of the values of its k nearest neighbour. It can be useful to weight the contributions of the neighbour, so that the nearer neighbour contribute more to the average than the more distant ones. The neighbour are taken from a set of objects for which the correct classification (or, in the case of regression, the value of the property) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. In order to identify neighbour, the objects are represented by position vectors in a multidimensional feature space. It is usual to use the Euclidian distance, though other distance measures, such as the Manhanttan distance could in principle be used instead.

2.1.4.3 *Bayesian networks*

A Bayesian network (BN) consists of a directed, acyclic graph and a probability distribution for each node in that graph given its immediate predecessors according to Su, Chopping, Rango, Martonchik and Peters (2007). A Bayes Network Classifier is based on a Bayesian network which represents a joint probability distribution over a set of categorical attributes. It consists of two parts, the directed acyclic graph G consisting of nodes and arcs and the conditional probability tables. The nodes represent attributes whereas the arcs indicate direct dependencies. The density of the arcs in a BN is one measure of its complexity. Sparse BNs can represent simple probabilistic models (e.g., naïve Bayes models and hidden Markov models), whereas dense BNs can capture highly complex models. Thus, BNs provide a flexible method for probabilistic modelling.

2.1.4.4 *Neural network*

An artificial neural network (ANN), often just called a "neural network" (NN), is a mathematical model or computational model based on biological neural networks, in other words, is an emulation of biological neural system. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Neural techniques have been used in feature extraction (Li, Wang and Tseng, 1998, Zhang 1996), stereo matching (Loung and Tan 1992) and image classification (Israel and Kasabov 1997).Filippi and Jensen (2006) are of the notion that artificial Neural Network (ANN) is appropriate for the analysis of nearly any kind of data regardless their statistical properties. ANN is very useful in extracting vegetation-type information in complex vegetation mapping problems, although it is at the expense of the interpretability of the results. ANN deploys a black-box approach which hides the underlying prediction process (Cerna and Chytry, 2005).

2.1.4.5. *Maximum likelihood classification*

Maximum-Likelihood Classifier (ML) and Mahalanobis Distance (MD) classifiers are usually regarded as the classic and most widely used supervised classification for satellite images resting on the statistical distribution pattern (Sohn and Rebello 2002; Xu et al. 2005). Both classifiers can show less satisfactory successes because their assumption that the data follow Gaussian distribution may not always be held in complex areas (Kapoor, Mehta, Esper, Poljsak-Prijatelj, Quan, Qaisar, Delwart and Li, 2010.The advantage of the MLC as a parametric classifier is that it takes into account the variance covariance within the class distributions and for normally distributed data, the MLC performs better than the other known parametric classifiers (Erdas, 1999). However, for data with a non-normal distribution, the results may be unsatisfactory. The first step in the classification process was the development of the classification scheme. The land cover classification scheme consisting of eight main land cover classes (mixed forest, degraded forest, herbaceous wetlands, shrub wetlands, grassland, grassland (open), mixed farmland and water (open)) they're developed based on the Afri-cover land cover classification system (Food and Agriculture Organization, 2005).

2.1.4.6 S*upport vector machine*

History of the support vector machines is ambiguous: it is difficult to name one single paper that introduced the concept. The fundamental theory of linear classifiers, which also includes support vector machines, dates back to the 1930's while the paper by Rosenblatt in 1956 introduced the perceptron of which the support vector machines are one very special case. Linear large margin classifiers which are the simplest form of the support vector machines have also been invented by several people in various fields of research, and similar ideas are presented in many articles, of which the paper by Vapnik and Lerner in 1963 and the paper by Mangasarian in 1965 are among. Generalization of linear large margin classifiers to the nonlinear case was introduced in Boser, Guyon and Vapnik,(1992), which is the first paper that presents the current support vector machine methodology, but due to earlier work it cannot be said that it is the essential paper that constitutes the principles of the support vector machines.

Support vector machines (SVM) are a group of relatively novel statistical learning algorithms that have not been extensively exploited in the geospatial science community, (Shi and Yang, 2012). Their basic idea is to construct separating hyperplanes between classes in feature space through the use of support vectors which are lying at the edges of class domains. SVM seek the optimal hyperplane that can separate classes from each other with the maximum margin (Vapnik, 1995). SVM were originally designed as a binary linear classifier, which assumes two linearly separable classes to be partitioned. In most cases, the best separable hyperplane may not be located exactly between two classes. To account for this, an error item is introduced to manipulate the tradeoff between maximizing the separation margin and minimizing the count of training samples that locates on the wrong side. SVM are further extended to deal with non-linear classification by using a non-linear kernel function to replace the inner product of optimal hyperplane. Several commonly used kernel functions include linear kernel, polynomial kernel, radial basis function (BRF), and sigmoid kernel (Haykin, 1999). Each of these kernel functions is constructed with multiple parameters, and how these parameters are setting can influence the performance of a specific support vector machine (Yang, 2011). However, Support Vector Machines (SVMs) have been recently proposed as a new technique for pattern recognition. Intuitively, if given a set of points which belong to either of two classes, a linear SVM finds the hyperplane leaving the largest possible fraction of points of the same class on the same side, while maximizing the distance of either class from the hyperplane, the hyperplane will then be determined by a subset of the points of the two classes, named support vectors, and has a number of interesting theoretical properties (Pontil and Verri, 1998).

In machine learning, support vector machines (SVMs, also support vector networks are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

Theoretical properties of the support vector machines have been studied in conjunction with general research on the theoretical properties of machine learning. Theory on which the support vector machines are based is a new generation of machine learning algorithms that take their inspiration from statistical learning theory (Gunn, 1998) and it has changed over time. The first explanation was given by the Vapnik-Chervonenkis theory1, followed by the theory of large margin classifiers, but these explanations have still left something to hope for. The latest theory, called data dependent structural risk minimization, and, in particular, the luckiness framework, appears to remedy the lack of theory, and it also combines the VC-theory and some parts of the theory of large margin classifiers within the framework of one single theory. It should be pointed out that in computer science, the VC-theory and parts of the theory of large margin classifiers is called PAC-theory (Probably Approximately Correct). The concept of this principle is to search a hypothesis h for which we can assure the lowest true error. Such error of h is the probability that h will make an error on an unseen and arbitrarily selected test example. A maximum limit can be used to link the true error of a hypothesis h with the error of h on the training set and the difficulty of H (measured by VC-Dimension), the hypothesis space having h. Support vector machines find the hypothesis h which (nearly) reduces this limit on the true error by properly handling the VC Dimension of H.

SVMs aim at minimizing an upper bound of the generalization error through maximizing the margin between the separating hyperplane and the data. This can be regarded as an approximate implementation of the Structural Risk Minimization (SRM) principle, which endows SVMs with good generalization performances independent of underlying distributions (Joachims, 1999). SVMs are a system for efficiently training the linear learning machines in the kernel- induced feature space (Cristianini and Shawe-Taylor, 2000). SVMs were originally designed as a linear classifier, but they are easily extended to nonlinear classifiers by mapping the space $S = \{x\}$ of the input data into a high-dimensional feature space $F = \{\Phi(x)\}$. By choosing an adequate mapping $\Phi(x)$, the data points that cannot be linearly separated in the input space become linearly separable or mostly linearly separable in the high-dimensional feature space, so that one can easily apply the structure risk minimization.

Compared with the traditional way of implementing mapping functions, SVMs have incomparable advantages. We need not compute the mapped patterns $\Phi(x)$ explicitly, and instead we only need the dot products between mapped patterns. They are directly available from the kernel function which generates $\Phi(x)$ Amari and Wu, 1999). SVMs are an elegant and highly principled learning method for classifying nonlinear input data. SVMs are gaining increasing popularity due to a number of attractive features including (Cristianini and Shawe-Taylor, 2000):

(i). Characteristics of support vector machine

1. SVMs are statistics-based models rather than loose analogies with natural learning systems. SVMs are theoretically related to a wide variety of study fields related to regularization theory and sparse approximation.

2. SVMs do not incorporate problem-domain knowledge. No assumption for input data distribution or error structure is needed in SVMs.

3. SVMs have a promising generalization performance. The formulation embodies the structural risk minimization (SRM) principle, as opposed to the empirical risk minimization (ERM) approach commonly employed within statistical learning methods. SVMs provide a method for controlling model complexity independently of dimensionality. It is this difference which equips SVMs with an excellent potential to generalize.

4. SVMs have the ability to model non-linear relationships in an effective and efficient way. SVMs operate in a kernel induced feature space. By using a suitable inner-product kernel, SVMs allow for constructing non-linear classifiers using only linear algorithms.

5. SVMs have the property of condensing information in the training data and providing a sparse representation by using a very small number of data points, namely, support vectors (SVs). Therefore, computations can be performed efficiently. This is especially true for huge datasets.

6. SVMs can guarantee a global and in general unique optimum. SVMs use quadratic programming to achieve maximized margin separation, which provides global minima only. The absence of local minima is a significant difference from the neural network classifiers.

7. SVMs have an extra advantage regarding automatic model selection in the sense that both the optimal number and locations of the support vectors are automatically obtained during training (Schölkopf, Burges and Smola, 1999).

8. SVMs are a robust tool for classification and regression in noisy, complex domains.

Owing to these prominent features, SVMs have gained growing popularity in recent years and have been successfully applied to a variety of fields such as text categorization, image recognition, hand-written digit recognition, potential disease spread prediction (Guo, Kelly and Graham, 2005) and land cover classification (Huang, Davis, and Townshed, 2002).

(ii)Extended SVMS

SVMs have been extended to meet the requirements of different applications. Some important extensions of SVMs are:

One-class SVMs (Schölkopf et al., 2001): One extension of SVMs to handle one-class classification problem in which only the training data of one class is available and the target class is modeled by fitting a hypersphere with minimal radius around it. Multi-class SVMs:

SVMs are extended to deal with K-class pattern classification problem. Multi-class SVMs are usually implemented by combining several binary SVMs to solve a given multi-class problem. Popular methods are: one-versus-all method using winner-takes-all strategy (Hastie and Tibshirani, 1998), which constructs K hyperplanes between class k and the K −1 other classes; and one-versus-one method implemented by max-wins voting (Platt, 1999), which Support Vector Regression (SVR) (Smola, 1996): One extension of SVMs to apply to regression task by the introduction of an alternative loss function ε intensive loss.

Reduced SVMs (Lee and Mangasarian, 2000): The reduced support vector machine (RSVM) is proposed to avoid the computational difficulties in classifying massive dataset by selecting a small random subset from the entire dataset to generate a reduced kernel (rectangular) matrix without sacrificing the prediction accuracy.

Least Squares SVMs (Suykens and Vandewalle, 1999): Least squares support vector machines (LS-SVMs) are re-formulations to the standard SVMs by using a regularized least squares cost function with equality constraints, leading to linear Karush-Kuhn-Tucker systems. The solution can be solved efficiently by iterative methods like the conjugate gradient algorithm.

(iii) Advantages of SVM over other machine learning algorithms

A major advantage over many other methods is that the SVM has a global solution, while other methods usually have many local solutions. Bekkari, Eldbraim, Elhassony, Mammass, El yassa and Dcrot,(2012, called SVM a group of advanced machine learning algorithms that have seen increased use in land cover studies. One of the theoretical advantages of the SVM over other algorithms (decision trees and neural networks) is that it is designed to search for an optimal solution to a classification problem whereas decision trees and neural networks are designed to find a solution, which may or may not be optimal. This theoretical advantage has been demonstrated in a number of studies where, SVM generally produced more accurate results than decision trees and neural networks.

SVMs have been used recently to map urban areas at different scales with different remotely sensed data. High or medium spatial resolution images (e.g., IKONOS, Quickbird, Landsat (TM)/ (ETM+), SPOT) have been widely employed on urban land use classification for individual cities for; building extraction, road extraction and other man-made objects extraction.

2.1.5 Support Vector Machine and Data Mining Classification

Support vector machines (SVMs) have been promising methods for classification and regression analysis because of their solid mathematical foundations which convey several salient properties that other methods hardly provide. However, despite the prominent properties of SVMs, they are not as favoured for large-scale data mining as for pattern recognition or machine learning because the training complexity of SVMs is highly dependent on the size of a data set. Many real-world data mining applications involve millions or billions of data records where even multiple scans of the entire data are too expensive to perform.

Data mining is a process of extraction of hidden knowledge, exceptional patterns and new findings from huge databases. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data pattern analysis (Jiawei, Han and Micheline, Kamber; 2012). Classification is one of the major roles in Data mining. Basically classification is a 2-step process; the first step is supervised learning for the sake of the predefined class label for training data set. Second step is classification accuracy evaluation. The significance of the support vectors in support vector machine (SVM) classification is intended to minimise confusion between classes (Huang et al., 2002). One advantage of using support vector machine is its extension from two classes to include multiclass classification. This is done by adopting a multiclass approach (Vapnik, 1998). Several advanced approaches have been proposed and used in multiclass classification. One of approach is one-against-one (Melgani & Bruzzone, 2004; Vapnik, 1998). The use of one-against-one approach helps in building more classes and it keeps the size of training data smaller for training (Melgani & Bruzzone, 2004).

2.1.6 Application of Support Vector Machine in Remote Sensing

Plaza et al. (2009), focused on recent developments in methodologies for processing a specific type of imagery, for example hyperspectral images. In its simplest form, SVMs are linear binary classifiers that assign a given test sample a class from one of the two possible labels. An instance of a data sample to be labelled in the case of remote sensing classification is normally the individual pixel derived from the multi-spectral or hyperspectral image. Such a pixel is represented as a pattern vector, and for each image band, it consists of a set of numerical measurements. Elements of the feature vector may also include other discriminative variable measurements based on pixel spatial relationships such as texture. An important generalization aspect of SVMs is that frequently not all the available training examples are used in the description and specification of the separating hyperplane. The subset of points that lie on the margin (called support vectors) are the only ones that define the hyperplane of maximum margin.

The implementation of a linear SVM assumes that the multi-spectral feature data are linearly separable in the input space. In practice, data points of different class memberships (clusters) overlap one another. This makes linear separability difficult as the basic linear decision boundaries are often not sufficient to classify patterns with high accuracy. Although, the classification accuracy produced by support vector machine (SVM) depends on the type of kernel function used, Gaussian radial basis filter (RBF) kernel is the most widely applied kernel function in support vector machine (SVM) classification (Foody & Mathur, 2004a & 2004b). This is because the support vectors that are used in the classification are controlled by the kernel specific function parameter through cross validation (Vapnik, 1995).

Furthermore, typical remote sensing pixels were first shuffled, with each image instance suffering the same random permutation. Yet, when the act of 'vandalism' (or removal of prior knowledge) took place, SVM still outperformed even the best neural networks. This discovery is particularly appealing in remote sensing applications since data acquired from remotely sensed imagery usually have unknown distributions, and methods such as Maximum Likelihood Estimation (MLE) that assume a multivariate normal data model do not necessarily match that assumption. Even if the data, whose dimensionality is assumed to match the number of spectral bands, were normally distributed, the assumption that the distribution can be described using a bell-shaped (Gaussian) function ceases to be sound, since the concentration of data in higher dimensional space tends to be in the tails (Fauvel et al., 2009). This phenomenon will continue to be encountered in remote sensing as new sensors increase spectral resolution and therefore data dimensionality.

2.1.7. Statistical Learning Theory

Support vector machines are based on a principle from computational learning theory which is called as structural risk minimization principle. To describe the idea of SVMs, the issue of structural risk minimization principle has to be addressed first. Therefore, we will start with posing a generic statistical learning problem.

2.1.7.1. *Statistical learning problem*

Consider a binary classification problem: suppose we are given empirical observations (thereafter called training set),

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_m, y_m) \in \mathbf{X} \times \mathbf{Y}, \mathbf{X} = \mathbf{R}^n, \mathbf{Y} = \{-1,+1\} \tag{2.1}$$

where **X** is the input space of potential observations, and **Y** is the possible decision space. Assume that the training set is drawn independently from an unknown (but fixed) probability distribution *P* (**X, Y**). This is a standard assumption in learning theory. Data generated this way is commonly referred to as IID (independent and identically distributed). The goal of classification problem is to find a classifier y = $f(\mathbf{x})$, which is a map from **X** to **Y** based on data in **T**. Any future case (outside training set **T**) that is also generated from *P* (**X, Y**) will be classified correctly by the map found. Of course, no classifier can classify every unseen example perfectly. Correctness of the classification is then measured by ***a loss function*** *L* (x, y, *f* (x)).

2.1.7.2 ***Loss function***

Loss Function is denoted by ($\mathbf{x}$, y, f ($\mathbf{x}$)): $\mathbf{x}\in\mathbf{X}$, $y\in\mathbf{Y}$, $f(\mathbf{x}) \in\mathbf{Y}$ the triplet consisting of a pattern $\mathbf{x}$, anobservation y and a prediction $f(\mathbf{x})$. Then the map L: $\mathbf{X} \times \mathbf{Y} \times \mathbf{Y} \rightarrow [0, \infty)$ with the property $L(\mathbf{x}, y, y) = 0$ for all $\mathbf{x}\in\mathbf{X}$ and $y\in\mathbf{Y}$ will be called a loss function.

The well-known loss functions are squared loss used in least square algorithm and logistic loss used in logistic regression.

$$\text{Squared loss}: L(\mathbf{x}, y, f(\mathbf{x})) = (f(\mathbf{x}) - y)^2 \tag{2.2}$$

$$\text{Logistic loss}: L(\mathbf{x}, y, f(\mathbf{x})) = \ln(1 + e^{-y \cdot f(\mathbf{x})}) \tag{2.3}$$

The former uses the square of the amount of mis-prediction to determine the quality of the estimate. It satisfies the assumption that we have additive normal noise corrupting the observations. The latter uses the product $yf(\mathbf{x})$to assess the quality of the estimate, where the sign of the prediction sgn (f (x)) denotes the class label, and the absolute value $|f(\mathbf{x})|$ describes the confidence of the prediction. No penalty occurs if $yf(\mathbf{x})$ is sufficiently large, i.e. if the patterns are classified correctly with large confidence. The logistic loss is used in order to associate a probabilistic meaning with prediction $f(\mathbf{x})$.

In a binary classification problem, another kind of loss function, namely, zero-one loss function is generally used:

$$L(\mathbf{x}, y, f(\mathbf{x})) = \frac{1}{2}\left| f(\mathbf{x}) - y \right| \qquad (2.4)$$

Note that the loss is 0 if the sample ($\mathbf{x}$, y) is classified correctly and 1 otherwise.

2.1.7.3 *Risk function*

To sum up the total expected loss for any mapping f: X × $\mathbf{A} \rightarrow \mathbf{Y}$, where A is the parameter space for the mapping function, a risk function comprising the loss and the underlying probability distribution is used:

$$R(\alpha) = \int L(\mathbf{x}, y, f(\mathbf{x}, \alpha))dP(\mathbf{x}, y) = \int \frac{1}{2}\left| f(\mathbf{x}, \alpha) - y \right| dP(\mathbf{x}, y) \qquad (2.5)$$

Where $f(\mathbf{x}, \alpha)$ is a classifier from a fixed parametric family $\{f(\mathbf{x}, \alpha): \alpha \in \mathbf{A}\}$.

Any choice of a particular α produces a classifier. The goal of statistical learning is to find a classifier with the minimal expected risk (or simply called risk). The difficulty of the task stems from the fact that we are trying to minimize a quantity that we cannot actually evaluate: since the underlying probability distribution P (**X, Y**) is usually unknown, it is impossible to compute the integral (2.5) and thus to achieve the risk minimization directly.

We do not know the probability distribution P (**X, Y**) that potential observations will be generated from. We do know, however, the training set **T** is generated from P (**X, Y**). Thus, we can try to infer a classifier $f(\mathbf{x}, \alpha)$ from the training set that is, in some sense, close to the one minimizing the risk (2.5).

2.1.7.4 *Empirical risk minimization*

One way to proceed is to use the empirical distribution of the training set to approximate the underlying probability distribution P (**X, Y**) and thus to calculate an approximation for the integral in (2.5). This leads to the empirical risk:

$$\mathrm{R}_{emp}(\alpha) = \frac{1}{m}\sum_{i=1}^{m} L(\mathbf{x}_i, y_i, f(\mathbf{x}_i, \alpha)) = \frac{1}{2m}\sum_{i=1}^{m}\left| y_i - f(\mathbf{x}_{i,}\alpha) \right| \qquad (2.6)$$

Most traditional methods, e.g. least square estimate, maximum likelihood estimate, and artificial neural network, aim to achieve empirical risk minimization. This makes some sense since according to the theory of uniform convergence in probability:

$$\lim_{m \to \infty} p\left\{ \sup_{\alpha \in \mathbf{A}} (R(\alpha) - R_{emp}(\alpha)) > \varepsilon \right\} = 0, \forall \varepsilon > 0 \qquad (2.7)$$

The empirical risk will infinitely approximate the expected risk when the size of training set increases. However, the size of the training set is limited. We are not sure how well the empirical distribution of the training set can approximate the unknown probability distribution P (**X, Y**). Therefore, minimizing the empirical risk does not always imply a small expected risk. For example, consider the 1D classification problem shown in Figure 2.4, with a training set of three points (marked by circles), and three test inputs (marked on the x-axis). Classification is performed by thresholding real-valued functions $g(x)$ according to sgn($g(x)$). Note that both classifiers represented using dotted line and solid line can perfectly explain the training data, but they give opposite predictions on the test inputs. That is, both classifiers can achieve empirical risk minimization but lead to quite different expected risks. Lacking any further information, the training data alone provides no means to tell which of the two functions is to be preferred.

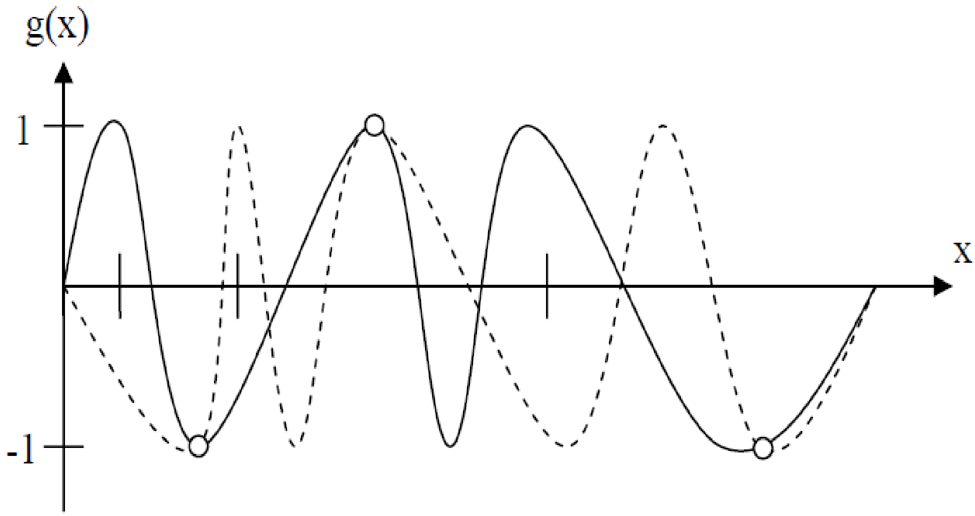


Figure 2.4: A 1D classification problem

Source: Chenglin Xie, 2006

2.1.7.5 *Structural risk minimization*

Although we cannot calculate the expected risk if the underlying probability is unknown, it is possible to find an upper bound for the expected risk and pose a problem for its minimization. Instead of empirical risk minimization, statistical learning theory (or Vapnik-Chervonenkis theory) aims to find a learning machine (classifier) with the minimum upper bound on the expected risk. This leads us to a method of choosing an optimal classifier for a given task. This is the essential idea of the structural risk minimization.

Prior to discussing SRM, we need to introduce the notion of function set capacity and define some means of measuring that capacity. The term **capacity** can be introduced as the ability of a machine (a parametric family or function set) to learn any training set without an error. Suppose we have a training set of m samples that can be assigned labels +1 or -1. Clearly, there are 2^m ways to label the training set. If, for each labeling, there is a classifier in the function set {f ($\mathbf{x}$, α): $\alpha \in \mathbf{A}$} that can correctly assign those labels, we say that the training set can be shattered by the function set. Maximum cardinality of the training set that can be shattered by {f ($\mathbf{x}$, α): $\alpha \in \mathbf{A}$} is called the Vapnik-Chervonenkis (VC) dimension of that function set. VC-dimension is clearly a property of the parametric family and can then be used as a measure of capacity of a particular learning machine belonging to that family (Vapnik, 1995; Cristianini and Shawe-Taylor, 2000).

The VC dimension sounds a little abstract. A simple example might be helpful in explaining it clearly(Figure 2.5). Considering a parametric family of hyperplanes in $\mathbf{R}^2$, as shown in Figure 2.6, it is capable of correctly classifying 3 samples with labels +1 or -1. There are 2^3 = 8 ways of assigning 3 samples to two classes. For the displayed samples in $\mathbf{R}^2$, all 8 possibilities can be realized using separating hyperplanes, in other words, the function class can shatter 3 samples. This would not work if we were given 4 points, no matter how we placed the hyperplanes. Therefore, the VC dimension of the class of separating hyperplanes in $\mathbf{R}^2$ is 3. It is easy to extend this conclusion to $\mathbf{R}^n$: the VC dimension of hyperplanes in $\mathbf{R}^n$ is $VCdim(H^n) = n + 1$.

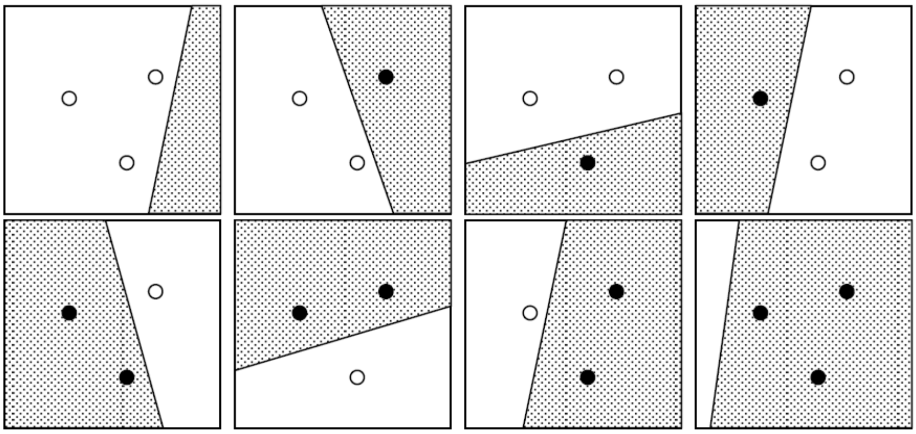


Figure 2.5: Anexample simple VC dimension

Source: Chenglin Xie, 2006.

Vapnik and Chervonenkis applied the probably approximately collect (PAC) model to statistical inference and gave the following theorem (Vapnik, 1995) to determine the upper bound for the expected risk.

Theorem 2.1: Supposing {f ($\mathbf{x}$, α): $\alpha \in \mathbf{A}$} is a parametric family for binary classification with adjustable parameters α. Then the following bound holds with a probability of at least1 − δ for any underlying distribution provided $h < m$:

$$R(\alpha) \le R_{emp}(\alpha) + \sqrt{\frac{h(\ln(2m/h)+1) - \ln(\delta/4)}{m}} \qquad (2.8)$$

where R (α) is the expected risk, R_{emp} (α) is the empirical risk, m is the size of the training set, and h is the VC dimension of the parametric family {f ($\mathbf{x}$, α): $\alpha \in \mathbf{A}$}.

The second term of (2.8) is called VC confidence interval. Given a training set of finite size, we can always come up with a learning machine which achieves zero training error (provided no examples contradict each other, i.e., whenever two patterns are identical, then they must come with the same label). To correctly separate all training examples, however, this machine will necessarily require a large VC *dimension h*. Therefore, the VC confidence interval, which increases monotonically with h, will be large. The bound (2.8) shows that the small training error does not guarantee a small test error. To achieve good generalization performance, that is, small expected risk, both the empirical risk and VC dimension of the parametric family have to be small. Figure 2.6 shows the relationships between VC dimension versus the empirical risk, VC confidence interval, and upper bound on the risk. Suppose we have a sequence of nested parametric families $S_1 \subset S_2 \subset \ldots \subset S_n \subset \ldots$ such that their VC dimensions satisfy $h_1 < h_2 < \ldots < h_n < \ldots$. With the increase of the VC dimension, it is possible to find a classifier in the parametric family to better fit the training set with finite size. Therefore, the empirical risk is usually a decreasing function of VC dimension h. As shown on (figure 2.6), the VC confidence interval will monotonously increase with the increase of VC dimension h. As a result, for a given size of training set, there is an optimal value of VC dimension which can achieve a minimal upper bound on the expected risk.

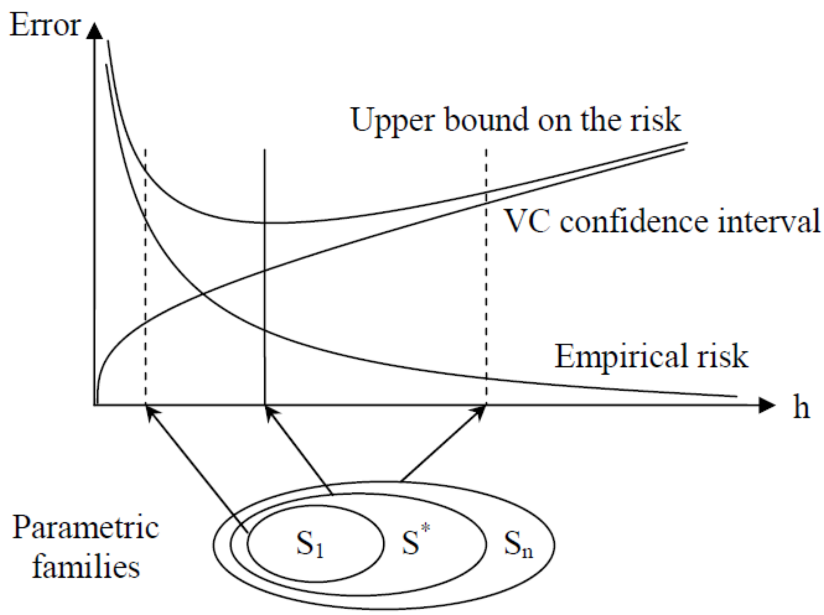


Figure 2.6: VC dimension vs. empirical risk, VC confidence, and the risk (Reprinted from Vapnik, 1995)

The choice of an appropriate VC dimension, which in some techniques is controlled by the number of free parameters of the model, is crucial in order to get good generalization performance, especially when the size of the training set is small. The objective of SRM then is to find an optimal VC dimension for which the upper bound on the expected risk is minimal. That is to minimize the empirical risk and VC confidence interval simultaneously, which can be achieved through the following two-stage process:

1. For each VC dimension h_i, identify a classifier $f(\mathbf{x}, \alpha)$ with minimal $R_{emp}(\alpha)$
2. In all the classifiers identified, choose the classifier, for which $R_{emp}(\alpha)$ + VC confidence interval is minimal.

However, finding a trade-off between reducing training error and limiting model complexity is not easy because the VC dimension of a parametric family can be hard to compute and there are only a small number of parametric families for which we know how to compute the VC dimension. Moreover, even if the VC dimension of a parametric family is known, it is difficult to solve the optimization problem of minimizing the empirical risk. Hence, we usually do not follow the above steps directly but rather use some more effective and efficient strategies. Support vector machines are able to achieve the goal of minimizing the upper bound of $R(\alpha)$ by efficiently minimizing a bound on the VC dimension h and $R_{emp}(\alpha)$ at the same time.

2.1.7.6. *Linear SVMs*

Support vector machines are an approximate implementation of structural risk minimization. Each particular choice of parametric families gives rise to a learning algorithm, consisting of performing SRM on the classifiers in parametric families of different VC dimensions. SVMs algorithms are based on parametric families of separating hyperplanes of different VC dimensions. SVMs can effectively and efficiently find the optimal VC dimension and an optimal hyperplane of that dimension simultaneously to minimize the upper bound of the expected risk.

Consider the problem of separating the training set of two separable classes of m examples:

$$\mathbf{T} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_m, y_m)\}, \mathbf{x} \in \mathbf{R}^n, y \in \{-1,1\} \tag{2.9}$$

with a hyperplane parameterized by **w** and *b*, $(\mathbf{w}, b) \in \mathbf{R}^n \times \mathbf{R}$,

$$\mathbf{w}' \cdot \mathbf{x} + b = 0 \tag{2.10}$$

where $\mathbf{x}_i$ is a data point in n-dimensional space, y_i is a class label, **w** is n-dimensional coefficient vector (w' is the transpose of **w**), and *b* is the offset. The discriminant function of the optimal hyperplane (classifier) is:

$$f(\mathbf{x}) = \mathrm{sgn}(\mathbf{w}' \cdot \mathbf{x} + b) \tag{2.11}$$

As shown in Figure 2.7, there exist many hyperplanes that can separate the examples perfectly. That is, many classifiers can achieve minimized empirical risk. Apparently, the generalization performances of these hyperplanes are quite different. Some hyperplanes, e.g. H_1 and H_4, have very poor generalization performance. Based on only the training set, how can we select a hyperplane which works well in general? According to structural risk minimization principle, we should select a hyperplane with a minimal VC confidence interval: that is, select a hyperplane with minimal VC dimension.

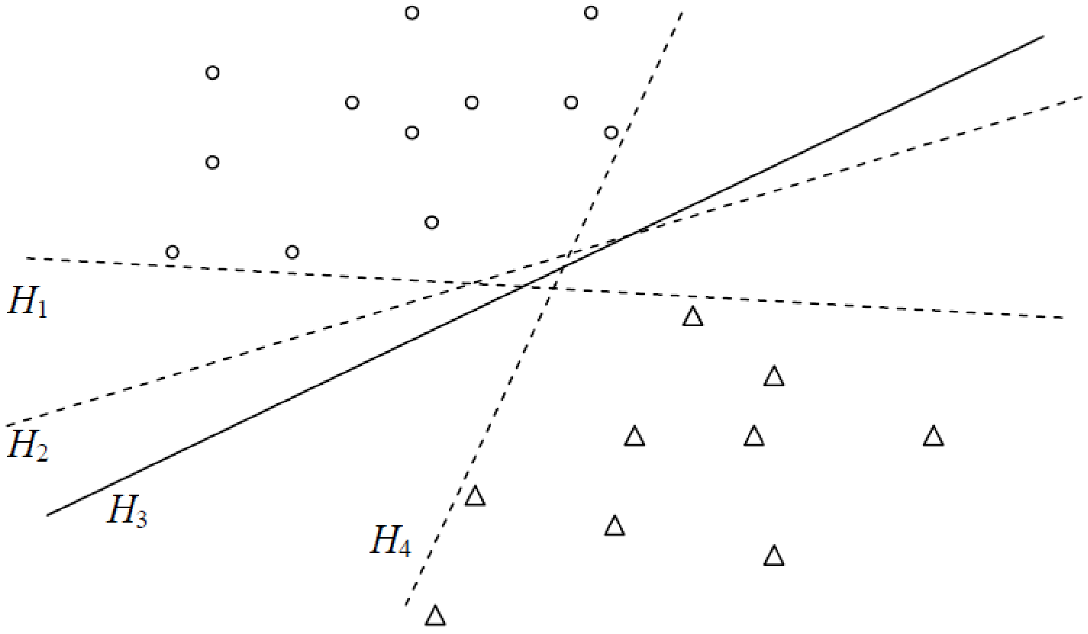


Figure 2.7: Hyperplanes perfectly separating two separable classes
Source: Chenglin Xie, 2006.

Vapnik (1995) formulated another theorem to determine a separating hyperplane with the minimal VC dimension.

Theorem 2.2: Let R be the radius of the smallest ball $B_R(\mathbf{a}) = \{\mathbf{x} \in \mathbf{T} : \|\mathbf{x} - \mathbf{a}\| \le R \,|\mathbf{a} \in \mathbf{T}\}$ containing the training set $\mathbf{T} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_m, y_m)\}, \mathbf{x} \in \mathbf{R}^n, y \in \{-1,1\}$, and let

$$f_{\mathbf{w},b}(\mathbf{x}) = \mathrm{sgn}(\mathbf{w}' \cdot \mathbf{x} + b) \tag{2.12}$$

be a canonical hyperplane decision function defined on the training set T. Then the set of separating hyperplanes $\{f_{\mathbf{w},b} : \|\mathbf{w}\| \le A\}$ has the VC dimension h bounded by

$$h \le \min(R^2 A^2, n) + 1 \tag{2.13}$$

According to theorem 2.2, in order to find a hyperplane with minimal VC dimension, we need to minimize the norm of the canonical hyperplane $\|\mathbf{w}\|$. A canonical separating hyperplane satisfied:

$$f_{\mathbf{w},b} : \min_{i=1,...,m} |\mathbf{w}' \cdot \mathbf{x}_i + b| = 1, \mathbf{x}_i \in \mathbf{T} \tag{2.14}$$

Specifically, we can find two hyperplanes parallel to the separating hyperplane and equal distances to it,

$$H_1 : y = \mathbf{w}' \cdot \mathbf{x}_i + b = +1 \qquad (2.15)$$

$$H_2 : y = \mathbf{w}' \cdot \mathbf{x}_i + b = -1 \qquad (2.16)$$

with the condition that there are no data points between H1and H 2. For any two parallel hyperplanes separating the data points, we can always scale the coefficient vector w and offset b so that they can be expressed as (2.15) and (2.16). As shown in Figure 2.8, the data points need to satisfy,

$$\mathbf{w}' \cdot \mathbf{x}_i + b \geq +1, \text{ for positive examples } y_i = +1, \quad i = 1,2,...,k \qquad (2.17)$$

$$\mathbf{w}' \cdot \mathbf{x}_i + b \leq -1, \text{ for negative examples } y_i = -1, \quad i = 1,2,...,k \qquad (2.18)$$

Conditions (2.17) and (2.18) can be combined into a single condition,

$$y_i(\mathbf{w}' \cdot \mathbf{x}_i + b) \geq 1 \qquad (2.19)$$

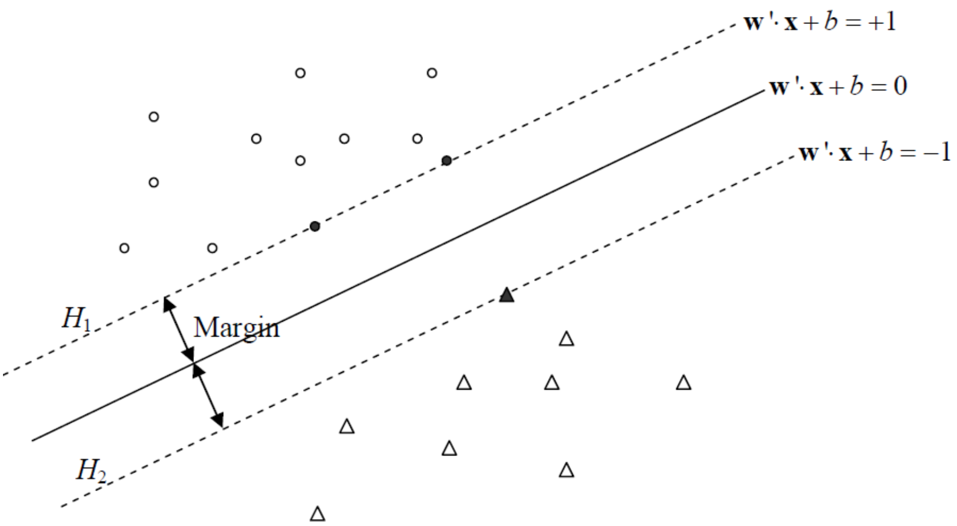


Figure 2.8: Optimal separating hyperplane between two classes of separable samples Source: Chenglin Xie, 2006.

The distance of a point $\mathbf{x}_i$from the hyperplane $\mathbf{w}' \cdot \mathbf{x}_i + b = 0$ is,

$$d(\mathbf{w},b;\mathbf{x}_i)=\frac{\left|\mathbf{w}^{'}\cdot\mathbf{x}_i+b\right|}{\|\mathbf{w}\|} \qquad 2.20$$

Therefore, the distance between H_1 and H_2 is given by,

$$\begin{aligned} d(H_1,H_2) &= \min_{\mathbf{x}_i:y_i=-1} d(\mathbf{w},b;\mathbf{x}_i)+\min_{\mathbf{x}_i:y_i=+1} d(\mathbf{w},b;\mathbf{x}_i) \\ &= \frac{1}{\|\mathbf{w}\|}\left(\min_{\mathbf{x}_i:y_i=-1}\left|\mathbf{w}^{'}\cdot\mathbf{x}_i+b\right|+\min_{\mathbf{x}_i:y_i=+1}\left|\mathbf{w}^{'}\cdot\mathbf{x}_i+b\right|\right) \qquad 2.21 \\ &= \frac{2}{\|\mathbf{w}\|} \end{aligned}$$

Consequently, minimizing the norm of the canonical hyperplane $\|\mathbf{w}\|$ is equal to maximizing the margin between H_1 and H_2. That is: the purpose of implementing SRM for constructing an optimal hyperplane is to find an optimal separating hyperplane that can separate the two classes of training data with maximum margin. Hence, there will be some positive examples on H_1 and some negative examples on H_2. Only these examples determine the optimal separating hyperplanes. Other examples have no contribution to the definition of the optimal separating hyperplanes and thus can be removed from the training set. Therefore, the examples located on H_1 and H_2 are called **support vectors**. The name ***Support Vector Machines*** originated from the name of support vector: that is, learning machines for finding support vectors.

Hence, the optimal hyperplane separating the training data of two separable classes is the hyperplane that satisfies,

$$\begin{aligned} &Minimize: F(\mathbf{w})=\frac{1}{2}\mathbf{w}^{'}\cdot\mathbf{w} \\ &subject\ \ to: y_i(\mathbf{w}^{'}\cdot\mathbf{x}_i+b)\geq 1, \qquad i=1,2,...,m \qquad (2.22) \end{aligned}$$

This is a convex, quadratic programming (QP) problem with linear inequality constraints. Problems of this kind are called constrained optimization problems. It is hard to solve the inequality constraint optimization problem directly. The most common way to deal with optimization problems with inequality constraints is to introduce Lagrange multipliers to convert the problem from the primal space to dual space and then solve the dual problem.

Introducing m non-negative Lagrange multipliers $\alpha_1, \alpha_2 \ldots, \alpha_m \geq 0$ associated with the inequality constraints in (3.22), we have the following Lagrangean function,

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^{'} \cdot \mathbf{w} - \sum_{i=1}^{m} \alpha_i [y_i(\mathbf{w}^{'} \cdot \mathbf{x}_i + b) - 1] \tag{2.23}$$

Solving the saddle point of Lagrangean function (which is unconstrained) is equivalent to solving the original constrained problem. At the saddle point of Lagrangean function, the gradient of $L(\mathbf{w}, b, \alpha)$ with respect to the primal variables **w** and b vanish,

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{m} \alpha_i y_i \mathbf{x}_i = 0 \tag{2.24}$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{m} \alpha_i y_i = 0 \tag{2.25}$$

Since these are equality constraints in the dual formulation, we can substitute them into $L(\mathbf{w}, b, \alpha)$ to yield,

$$L(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{i=1}^{m} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^{'} \cdot \mathbf{x}_j \tag{2.26}$$

Therefore, solving the constrained optimization problem (2.22) is converted to solving the dual optimization problem:

$$\begin{aligned}
&maximize: L(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{i=1}^{m} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^{'} \cdot \mathbf{x}_j \\
&subject\, to: \sum_{i=1}^{m} \alpha_i y_i = 0 \\
&\qquad\qquad \alpha_i \geq 0, \qquad i = 1, 2, \ldots, m
\end{aligned} \tag{2.27}$$

This is a quadratic optimization problem. Over the years, a number of optimization techniques have been devised to solve the quadratic optimization problem. They range from the simple gradient ascent (the steepest ascent) algorithm to more efficient algorithms, namely, the Newton method, conjugate gradient method, and primal dual interior-point method. These methods can be applied in SVMs to solve the above mentioned optimization problem. However, many of these methods require that the matrix $y_i\, y_j \mathbf{x}_i' \cdot \mathbf{x}_j$ is stored in memory. This implies that the space complexity of the algorithm is quadratic in the sample size. For large size problems, these approaches can be inefficient and sometimes impossible.

A novel algorithm called Sequential Minimal Optimization (SMO) (Platt, 1998) was designed to solve large size quadratic optimization problems. The strategy of SMO is to decompose the problem into a series of small tasks that optimize a minimal subset of just two variables at each step. An analytical solution for the two variables optimization problem is given and the original problem can be solved using iteration.
After obtaining an optimal solution $\alpha^* = (\alpha_1^*, \alpha_2^*, ..., \alpha_m^*)$ for the dual problem, the solution of an optimal coefficient vector for the primal problem can be obtained from (2.24):

$$\mathbf{w}^* = \sum_{i=1}^{m} y_i \alpha_i^* \mathbf{x}_j \tag{2.28}$$

The offset *b* does not appear in the QP problem and the optimal solution *b** must be solved from the primal constraints,

$$b^* = -\frac{1}{2}(\min_{\mathbf{x}_i : y_i = +1} \mathbf{w}^{'} \cdot \mathbf{x}_i - \max_{\mathbf{x}_i : y_i = -1} \mathbf{w}^{'} \cdot \mathbf{x}_i) \tag{2.29}$$

Generally, we do not need to calculate **w*** explicitly. A new example x can be classified using:

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w}^{*'} \cdot \mathbf{x} + b^*) = \text{sgn}((\sum_{i=1}^{m} \alpha_i^* y_i \mathbf{x}_i)^{'} \cdot \mathbf{x} + b^*) = \text{sgn}(\sum_{i=1}^{m} \alpha_i^* y_i \mathbf{x}_i^{'} \cdot \mathbf{x} + b^*) \tag{2.30}$$

Karush-Kuhn-Tucker (KKT) complementarity conditions (Taha, 1997) of optimization theory require that:

$$\alpha_i^*[y_i(\mathbf{w}^{*'} \cdot \mathbf{x}_i + b^*) - 1] = 0, \qquad i = 1,2,...,m \tag{2.31}$$

Therefore, only examples $\mathbf{x}_i$ that satisfy the equalities in (3.19) can have non-zero coefficients α_i^*. Such examples lie on the two parallel hyperplanes separating two classes and thus are support vectors. Therefore, support vectors are examples whose related coefficients α_i^* are non-zero.

Since only a small part of examples are located on the two parallel hyperplanes, most examples satisfy the inequalities in (3.19), i.e., most α_i^* solved from the dual problem are null. Therefore, the coefficient vector w is a linear combination of a relatively small percentage of examples (support vectors). This leads to a sparse solution and it is very efficient in classifying new examples. Since only support vectors have non-zero coefficients α_i^*. The new example can be classified according to only support vectors,

$$f(\mathbf{x}) = \mathrm{sgn}\Big(\sum_{\text{support vetor}} \alpha_i^* y_i \mathbf{x}_i^{'} \cdot \mathbf{x} + b^*\Big) \tag{2.32}$$

2.1.7.7. *Soft Margin SVMs*

In practice, not all training sets can be perfectly linearly separated by a hyperplane (Figure 2.9). In the case that the training set T is not linearly separable or we want to consider a general case and simply ignore whether or not the set T is linearly separable, the algorithm discussed earlier needs to be extended to solve imperfect separation problems. In that case, SVMs do not strictly require that there are no examples between separating hyperplanes H_1 and H_2. Instead, a penalty for the examples that cross the boundaries is introduced to take into account the misclassification errors.

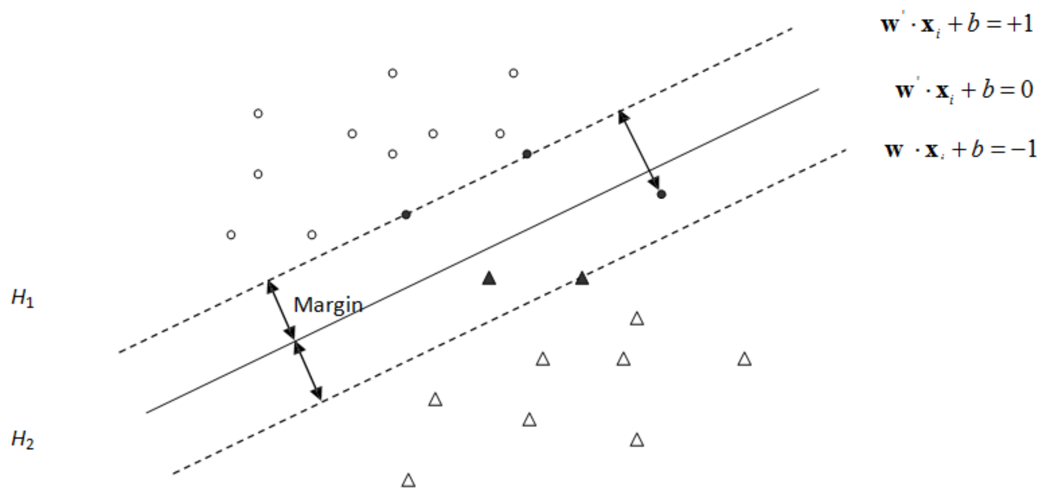


Figure 2.9: Optimal separating hyperplane between two classes of inseparable samples Source: Chenglin Xie, 2006.

This makes sense from the structural risk minimization point of view. A good generalization performance can be reached when both the empirical risk and the VC confidence interval are small. Therefore, minimizing the VC dimension of the classifier by maximizing the margin can lead to a minimal VC confidence and thus a minimal upper bound for the expected risk given that (2.19) has to be met. Then, the question arises: is it possible to allow for a small number of misclassified points in order to achieve better generalization performance?

The answer is quite obvious. If the decrease in the VC confidence interval caused by a simpler classifier is larger than the increase of empirical risk caused by the misclassification error of imperfect separation, the upper bound of the expected risk will decrease and thus lead to a good generalization performance. Actually, this is a generalized optimal separating hyperplane (OSH) problem. Soft margin SVMs aim to minimize the upper bound of the expected risk by minimizing the trade-off between margin (VC dimension, or VC confidence interval) and training error (empirical risk). To handle imperfect separation problems, non-negative slack variables ξ_i are incorporated into constraints (2.19) to consider misclassification errors:

$$y_i(\mathbf{w}' \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \qquad i = 1,2,...,m \tag{2.33}$$

$$\xi_i \geq 0 \qquad i = 1,2,...,m \tag{2.34}$$

Moreover, a penalty is added to the objective function to form a generalized expression for the upper bound of the expected risk,

$$F(\mathbf{w}, \xi) = \frac{1}{2}\mathbf{w}' \cdot \mathbf{w} + C(\sum_{i=1}^{m} \xi_i)^l \tag{2.35}$$

The first term in (2.35) is related to the VC dimension of the classifier and thus corresponds to the VC confidence interval in (3.8). The second term in (2.35) is related to the misclassified points for the training set and thus corresponds to the empirical risk in (2.8). The regularization parameter C is used to control the trade-off between the empirical risk and the model complexity. A large C corresponds to stronger penalties for errors and will lead to a complex model to minimize the number of misclassified points.

On the contrary, a small C corresponds to stronger penalties for complexity and will lead to a simple model to maximize the margin $\frac{2}{\|\mathbf{w}\|}$.

Usually, l is set to be 1. Hence the optimization problem becomes,

$$minimize: F(\mathbf{w},\xi) = \frac{1}{2}\mathbf{w}^{'} \cdot \mathbf{w} + C\sum_{i=1}^{m}\xi_i$$

$$subject\ to: y_i(\mathbf{w}^{'} \cdot \mathbf{x}_i + b) + \xi_i - 1 \geq 0, \quad i = 1,2,...,m \tag{2.36}$$

$$\xi_i \geq 0, \quad i = 1,2,...,m$$

Introducing Lagrange multipliers **α** and **β**, we have the following Lagrangean function,

$$L(\mathbf{w}, b, \xi, \alpha, \beta) = \frac{1}{2}\mathbf{w}^{'} \cdot \mathbf{w} + C\sum_{i=1}^{m}\xi_i - \sum \alpha[y_i(\mathbf{w}^{'} \cdot \mathbf{x}_i + b) + \xi_i - 1] - \sum_{i=1}^{m}\beta_i\xi_i \tag{2.37}$$

Similar to the linear separable case, we must now minimize L (**w**, b, ξ, **α**, **β**) with respect to **w**, b, ξ and simultaneously maximize L (**w**, b, ξ, **α**, **β**) with respect to **α**, **β**. Applying gradient vanishing conditions and simplifying yields,

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{m}\alpha_i y_i \mathbf{x}_i = 0 \tag{2.38}$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{m}\alpha_i y_i = 0 \tag{2.39}$$

$$\frac{\partial L}{\partial \xi} = C - \alpha_i - \beta_i = 0 \tag{2.40}$$

Substituting (2.38) -(2.40) into (2.37) yields the dual problem:

$$maximize: L(\alpha) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{i=1}^{m}\alpha_i\alpha_j y_i y_j \mathbf{x}_i^{'} \cdot \mathbf{x}_j$$

$$subject\,to: \sum_{i=1}^{m}\alpha_i y_i = 0$$

(2.41)

$$0 \leq \alpha_i \leq C, \quad i = 1,2,...,m$$

The only difference between perfectly separating case and the imperfectly separating case is that, the Lagrange multipliers α_i *s* are bounded above by C in an imperfectly separating case instead of unbounded in a perfectly separating case.

After the optimum Lagrange multipliers α_i have been determined, we can compute the optimum coefficient vector $\mathbf{w}^*$ and the optimal offset b^* . The solution is given by:

$$\mathbf{w}^* = \sum_{i=1}^{m} y_i \alpha_i^* \mathbf{x}_j \tag{2.42}$$

The offset b^* can be found from:

$$\alpha_i^*(y_i(\mathbf{w}^{*'} \cdot \mathbf{x}_i + b^*) - 1) = 0 \qquad (2.43)$$

For any *i* such that α_i^* is not zero.

Based on the new KKT conditions:

$$\alpha_i^*(y_i(\mathbf{w}^{*'} \cdot \mathbf{x}_i + b^*) - 1 + \xi_i^*) = 0 \qquad (2.44)$$

$$(C - \alpha_i^*)\xi_i^* = 0 \qquad (2.45)$$

The points in the training set can be classified into three categories:

1. $\alpha_i^* = 0$: normal points (non-support vectors)
2. $\alpha_i^* > 0$: support vectors
    a. $0 < \alpha_i^* <$ C : margin vectors
        - $\xi_i^* = 0$
        - The support vectors lie at a distance $\frac{1}{\|\mathbf{w}\|}$ from the OSH
    b. $\alpha_i^* = C$: non-negative vectors
        - $\xi_i^* > 1$: misclassified points
        - $0 \le \xi_i^* \le 1$: correctly classified within margin

Figure 2.10 visually shows these three kinds of points in the training set.

$\mathbf{w}^{'} \cdot \mathbf{x}_i + b = +1$

$\mathbf{w}^{'} \cdot \mathbf{x}_i + b = 0$

$\mathbf{w}^{'} \cdot \mathbf{x}_i + b = -1$

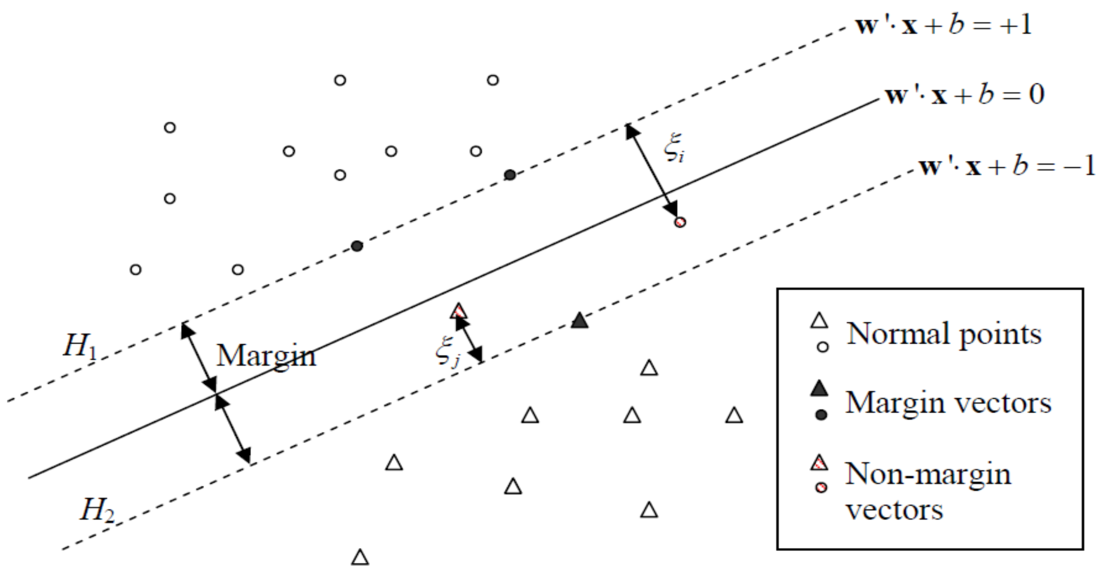


Figure 2.10: Three kinds of points in the training set

Source: Xie 2006

2.1.7.8. *Non-Linear SVMs*

In most practical applications, the two classes cannot be linearly separated. To extend the linear learning machine to work with non-linear cases, SVMs use a kernel method to map the non-linearly separable classes from input space to a high dimensional feature space, in whichthe non-linearly separable classes can be separated by a linear optimal hyperplane.

As shown in Figure 2.11, a mapping function $\Phi(\mathbf{x})$ (quadratic transform) is used to map examples in the input space $\mathbf{S}\subset\mathbf{R}^2$ into a high dimensional feature space $\mathbf{F}\subset\mathbf{R}^3$. The training set, which cannot be linearly separated in the input space, now become linearly separable in the feature space.

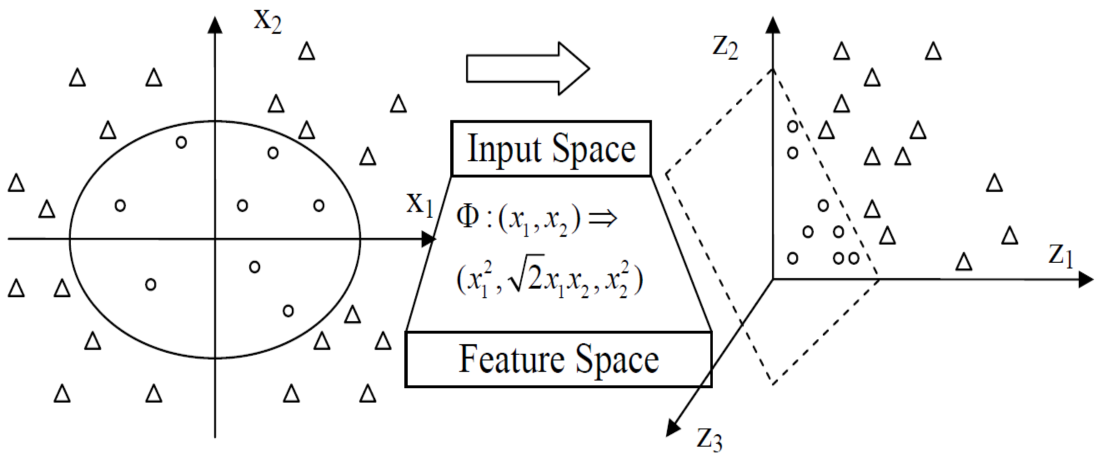


Figure 2.11: Mapping from input space to feature space (reprinted from Smola et al., 1999)

From the above example, we can find that appropriate choice of mapping function $\Phi(\mathbf{x})$, that is, appropriate construction of feature space, leads to linear separability. However, explicit use of such a mapping function would cause some efficiency problems. The dimension of the feature space is usually much higher than that of the input space, which will dramatically increase the number of parameters need to be solved. For example, polynomial transformation of degree d over N attributes in the input space leads to $\binom{d+N-1}{d} = \frac{(d+N-1)!}{d!(N-1)!}$ attributes in the feature space. Moreover, the transformation operator $\Phi(\mathbf{x})$ might be computationally expensive.

After checking the optimal separating hyperplane problem, however, it is easy to find that: an example x in the input space can be represented as Φ(x) in the feature space. Since the linear separation is performed in the feature space, the optimization problem shown in (2.41) can be rewritten as:

$$maximize : L(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{i=1}^{m} \alpha_i \alpha_j y_i y_j \Phi(\mathbf{x}_i') \cdot \Phi(\mathbf{x}_j) \tag{2.46}$$

Since only the dot product of two vectors in the feature space appears in the optimization problem, we can define a kernel function K as,

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_j)' \cdot \Phi(\mathbf{x}_j) \tag{2.47}$$

Hence, we do not need to know explicitly the mapping function, but can simply use a ***kernel function*** (KF) of the input space to represent the dot product in the high dimensional feature space. All the previous derivations in linear SVMs hold (substituting dot product with the kernel function), since we are still doing a linear separation, but in a different space.

The use of the kernel function greatly simplifies the mapping problem and improves the computational efficiency. While the mapping function needs to map $\mathbf{R}^n$ to $\mathbf{R}^l$ (usually n <<l), the kernel function can map $\mathbf{R}^n \times \mathbf{R}^\mathrm{n}$ to $\mathbf{R}$ and thus reduce the computational burden dramatically. For example, consider the following map:

$$\Phi(\mathbf{x}) = (x_1^2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, 1) \tag{2.48}$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_j)'\cdot\Phi(\mathbf{x}_j) = (\mathbf{x}_i'\cdot\mathbf{x}_j + 1)^2 \tag{2.49}$$

Using the kernel function, we can greatly simplify the computation. The kernel (3.49) is known as second order polynomial kernel.

Hence the optimization problem can be rewritten as,

$$maximize: L(\alpha) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{i=1}^{m}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{2.50}$$

Thus the new example can be classified according to,

$$f(\mathbf{x}) = sign(\sum_{\text{support vetor}} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b^*) \tag{2.51}$$

The existence of a kernel function and an appropriate feature space is problem-dependent and has to be established for each new problem. The following lists some commonly used kernels:

- Dot kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i'\cdot\mathbf{x}_j$
- Polynomial kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i'\cdot\mathbf{x}_j + 1)^d$
- Radial basis kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-|\mathbf{x}_i'\cdot\mathbf{x}_j|^2/\sigma^2}$
- Sigmoid kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma\,\mathbf{x}_i'\cdot\mathbf{x}_j + c)$

According to the definition of kernel function (2.47), it should be able to be expressed as dot product in a high dimensional space. According to Mercer's condition (Cristianini and Shawe-Taylor, 2000), any positive definite function $K(x, y)$ can be expressed as a dot product in a high dimensional space. Therefore, any kernels that meet Mercer's condition can be used to construct SVMs:

$$\iint K(\mathbf{u}, \mathbf{v}) g(\mathbf{u}) g(\mathbf{v}) d\mathbf{u} d\mathbf{v} > 0, \qquad \forall g: \int g^2(\mathbf{u}) d\mathbf{u} < \infty \tag{2.52}$$

A kernel can also be constructed by combining other kernels. Assume K_1 and K_2 are kernels, then the following expressions can generate a valid new kernel:

- $K(\mathbf{x}_i, \mathbf{x}_j) = aK_1(\mathbf{x}_i, \mathbf{x}_j)$
- $K(\mathbf{x}_i, \mathbf{x}_j) = K_1(\mathbf{x}_i, \mathbf{x}_j) + K_2(\mathbf{x}_i, \mathbf{x}_j)$
- $K(\mathbf{x}_i, \mathbf{x}_j) = K_1(\mathbf{x}_i, \mathbf{x}_j)K_2(\mathbf{x}_i, \mathbf{x}_j)$
- $K(\mathbf{x}_i, \mathbf{x}_j) = e^{K_1(\mathbf{x}_i, \mathbf{x}_j)}$

2.1.7.9*The Three Main Mathematical Properties of SVMs*

The first property distinguishes SVMs from previous nonparametric techniques, like nearest-neighbors or neural networks. Typical pattern recognition methods are based on the minimization of the empirical risk that is on the attempt to minimize the misclassification errors on the training set. Instead, SVMs minimize the structural risk, which is the probability of misclassifying a previously unseen data point drawn randomly from a fixed but unknown probability distribution. If the VC-dimension of the family of decision surfaces is known, the theory of SVMs provides an upper bound for the probability of misclassification of the test set for any possible probability distributions of the data points (Vapnik and Chervonenkis, Vapnik, 1995).Second, SVMs condense all the information contained in the training set relevant to classification in the support vectors. This (a) reduces the size of the training set identifying the most important points, and (b) makes it possible to efficiently perform classification.Third, SVMs are quite naturally designed to perform classification in high dimensional spaces, even in the presence of a relatively small number of data points. The real limitation to the employment of SVMs in high dimensional spaces is computational efficiency. In practice, for each particular problem a trade-off between computational efficiency and success rate must be established.

2.1.8 Endmember

An endmember is defined as a spectrally pure pixel that portrays various mixed pixel in the image (Plaza et al., 2004b). The method of feature selection involves identifying the most discriminating measurements out of a set of D potentially useful measurements, where d ≤ D. Endmember extraction has been widely used in hyperspectral image analysis due to significantly improved spatial and spectral resolution provided by hyperspectral imaging sensor also known as imaging spectrometry (Chang et al., 2006; Chaudhry et al., 2006).

Identification of image endmember is a crucial task in hyperspectral data exploitation, especially classification (Martinez, Perez, Plaza, Anguilar, Cantero, and Plaza, 2006). When the endmembers have been selected, various methods can be used to construct their spartial distribution, associations and fractional abundances. For real hyperspectral data, various tools (algorithms) developed to execute the task of locating appropriate endmembers include, Pixel Purity Index (PPI), N-FINDR and Automatic Morphological Endmember Extraction (AMEE) (Plaza, Martinez, Perez, and Plaza, 2004a; Chang et al., 2006; Chaudhry et al., 2006;Martinez et al., 2006).

2.2Literature Review

Here is review of some research done on remote sensing and machine learning algorithm, presented in sections as general review, review of studies done in Nigeria (Local, Regional and National) and review of studies done outside Nigeria. They identified several areas for additional research, mostly related to appropriate treatments of some parametric and non-parametric factors in order to achieve improved mapping accuracies particularly for heterogeneous landscape types and for large datasets using various classifier and SVM in particular.

The general literature presented also reveals the gaps in the support vector machines technique to map various land cover types from a remote sensor image covering an urban area, and demonstrates the robustness of this geospatial technique for mapping heterogeneous landscapes. The review encompasses of studies done outside Nigeria and those of studies done within Nigeria both locally, regional and National.

2.2.1. General review

Moussa and El-Sheimy (2002) proposed an approach that can extracts Man-Made Objects and classification of same from Satellite/Aerial Imagery Using Neural Networks. Regions of these images were formed in vector format based on their colours as a preliminary step. The extracted regions were then analysed to form a set of descriptors to describe each region. The main harmonics of the boundaries of the extracted regions are transformed to form a reduced rotation-invariant representation that can be used as region descriptors. Also, the uniformity of the region texture was used as region descriptors. Their work uses a supervised neural network in which a set of inputs and associated outputs are provided to train the network at the initial stage. The training of the neural network uses the calculated descriptors for man-made objects from previously classified images. After the training phase, the descriptors of the test images objects were fed into the neural network to perform the classification. The results of the proposed approach are presented based on a reference solution for evaluation purposes. According to them no elevation profile exists for the regions, causing some classification ambiguities. The presented results for aerial and satellite images classification show the significance of the proposed technique in the classification process.

The used different feature types assisted their classification process. Almost, all the regular man-made regions were correctly classified, while the misclassified man-made objects are mainly irregular road segments affected by shadows.

The second test of the proposed approached was carried out using a QuickBird satellite image of an urban area. The region contains many shadow areas and many buildings with near-green colours. Based on manual classification of the image, 82% of the man-made objects are correctly classified, 23% of the natural objects are misclassified, and about 3% of the image is not classified due its very small size.The usage of different features that describes the shape, colour, and texture of man-made objects appears to assist the classification process. The high correct classification rate of the regular man-made regions shows the effectiveness of the presented shape features in the man-made classification. With the vast amount of available data for this problem, neural networks are good classifier candidate due to its ability of learning from similar data sets.

Gap: The proposed region-based technique allows further context features extraction that can help classify the shadows regions. The presented results for aerial and satellite images classification show the significance of the proposed technique in the classification process. However, the method used for the classification validation was not stated.

Sitaram and Manjunath (2006) suggested the use of texture motifs, or characteristic spatially recurrent patterns for modelling and detecting geospatial objects. The model is learned in a two-layered framework—the first teaches the constituent “texture elements” of the motif and the second, the spatial distribution of the elements.

In their experimental session, they first demonstrated the model training and selection methodology for different objects given a limited dataset of each. Then, emphasize the utility of such models for detecting the presence or absence of geospatial objects in large aerial image datasets comprising tens of thousands of image tiles.

Gap: The quality of the models, Gabor filter and Gaussian mixture model (GMM) was evaluated on the basis of their application to object detection. Experimental results demonstrate that such a modelling approach is quite effective in detecting complex geospatial objects. They illustrated the usefulness of their approach in reducing the manual labour involved in identifying object locations in large aerial image data sets. The combination of texture with other features, such as colour and shape increased the robustness of object detection.

Ktheirosh et al (2010) noted that automated detection of buildings in aerial data is important in many applications like map updating, city modelling, urban growth analysis and monitoring of informal settlements. Their methods were tested in two study areas. The first study area is a suburban neighbourhood located in the south of the city of Memmingen, Bavaria, Germany. This area comprises about seventy isolated buildings with dimensions ranging from 100 to 300m^2, many of which were gardens, garden sheds or garages. The second study area is an urban neighbourhood located at the centre of the German city of Mannheim. This area is characterized with large buildings, mostly attached forming building blocks of different heights, more cars and less vegetation.

However, they undertake a comparative evaluation of three common data fusion and classification methods, namely Bayesian, Dempster–Shafer and AdaBoost, as applied to the detection of buildings in multi-source aerial data. In addition, they compared the performance of the fusion methods with the less elaborate method of thresholding, the normalized DSM and presented results of both pixel-based and object-based implementations of the methods in an urban and a suburban study area, and compared the performance of the methods on the basis of ground truth information obtained by manual extraction of buildings. Their result shows that the method of thresholding a normalized DSM performs well in terms of the detection rate and reliability in the less vegetated Mannheim study area, but also yields a high rate of false positive errors. The Bayesian methods perform better in the Memmingen study area where buildings have more or less the same heights.

Gap: In both study areas, most of the errors were found at building boundaries and in areas where dense trees were present. In future research, they suggested that these errors will be investigated from three main aspects: (i) the usefulness of height texture descriptors as additional features for better classification of critical areas; (ii) the application of a second-level classifier (different from the first one) combined with rule-based approaches to remove ambiguity from the unclassified data resulting from the Dempster–Shafer and the AdaBoost methods; (iii) the comparison with more complex classifiers (such as SVM, Neural Networks and Particle- based approaches) over large amounts of data, and the method used for the classification validation was not stated,there is also need to use satellite remote sensing data of higher resolution to crosscheck this finding.

Bruzzone and Carlin (2006) carried out a multilevel context-based system for classification of very high spatial resolution images. They came up with a novel pixel-based system for the supervised classification of very high geometrical (spatial) resolution images. This system is aimed at obtaining accurate and reliable maps both by preserving the geometrical details in the images and by properly considering the spatial context information. The obtained results did not improve both the classification accuracies and the quality of the classification maps. This behaviour mainly depends on the criterion adopted for the definition of classes. Inasmuch as many classes share the same geometrical features (e.g., different kinds of building are discriminated only on the basis of the spectral signature), in this case, the use of the geometrical and relational information does not increase the separability.

Experiment 2, shows the comparisons with a feature-extraction module based on generalized Gaussian pyramid decomposition. The aim of the second set of experiments was to compare the proposed system with a different approach to multilevel. Feature extraction of very high resolution images based on the generalized Gaussian pyramid decomposition.The SVM classification module was also used in these trials. These results show that the proposed feature-extraction technique provided accuracy higher than the reference method.The accuracy obtained on test edge areas confirms the greater ability of the proposed approach (which increased Kappa values by 8.5% compared with the generalized Gaussian pyramid method) to model the geometrical details of objects in the scene, such as roofs and roads. A comparison between the accuracies obtained on homogeneous areas points out a gap of 2.4%.

Gap: It is worth noting that, unlike other approaches proposed in the literature and briefly described in this work, hierarchical segmentation does not aim to identify the best level of representation of each object, but simply models the multilevel spatial context of each pixel. This should be considered as a pre-processing stage aimed at driving the feature-extraction phase.In this case, the use of the geometrical and relational information does not increase the separability.

Bayoudh, Roux, Nock, and Gilles, (2012) carried outautomatic learning of structural knowledge from geographic information for updating land cover maps**.** Their study demonstrated the exploration of some strengths of artificial intelligence. Their main idea consists in inducing classification rules that explicitly take into account structural knowledge, using Aleph, an Inductive Logic Programming (ILP) system. They applied their proposed methodology to three land cover/use maps of the French Guiana littoral. One hundred and forty-six classification rules they are induced for the 39 land-cover classes of the maps. These rules are expressed in first order logic language which makes them intelligible and interpretable by non-experts. A ten-fold cross validation gave average values for classification accuracy, specificity and sensibility equal to, respectively, 98.82 %, 99.65% and 70%. The proposed methodology could be valuably exploited to automatically classify new objects and/or help operators using object-based classification procedures.

Their work involves implementing machine learning methods for structural and symbolic knowledge extraction from land use/cover maps and from various complementary geographic information layers. They define structural knowledge, on the one hand, as the knowledge that concern intrinsic structures of the land cover/use classes. They chose Inductive Logic Programming (ILP) to implement the machine learning thanks to clearness of the language used and the intelligibility of its resulting hypothesis.

They set the minimum accuracy of the candidate clauses to 0.7, accuracy being defined by p/(p+n) where p and n are respectively the number of positive examples and the number of negative examples covered by the clause, and the maximum clause length to 6 literals.

Gap: they implemented a method for automatic extraction of structural knowledge using Inductive Logic Programming and applied the proposed method to the updating of the land cover/use of the French Guiana littoral. Results show that the induced rules are intelligible and easy to interpret even by non-expert users. In particular, they provide knowledge on structural aspects. A ten-fold cross validation of their classifiers provided very promising results that suggest i) that the accuracy of the automatic classification procedures could be greatly improved by the addition of automatically learned structural knowledge and ii) that they are able to provide a valuable assistance to operators using object-based classification procedures.

Singh et al (2015)whoseresearch was on "Land Information Extraction with Boundary Preservation for High Resolution Satellite Image" had their study area located in Burghausen, Germany. They confirmed that extracting land information has been a challenging task for the modern world as there is no proper method available for this. They used manual digitization to extract land objects and used their database information for further work. They observed that this approach is time taking and with change in the land detail, updating the information is very challenging task. Object based techniques are the solution of this problem which was derived from the necessity of their use in the high resolution images. So their first aim was edge detection which can be used as one of the image layer during segmentation.

They developed an efficient technique for edge detection to define land boundaries and feature selection technique for land information extractionusing Rapid-Eye image with 5m resolution. So, they succeeded in using an edge detection technique and object based classification to extract the land information automatically and then associate the area detail with each land object.In this method they modify the Sobel operator, but this is not the solution for satellite images where they have effect of noise and texture variation.The highest gradient output and threshold value was selected using Otsu algorithm, (Lu and Weng, 2006).For better visual representation the raster boundary was converted into vector with the help of direct algorithm available in e-Cognition Developer.

Gap: The proposed method gives the alternative for information extraction from such images with preserving land object boundary by utilizing edge detection in which one can observe that traditional methods for edge detection didn’t shows much better results for satellite images as satellite images have complex texture and it doesn’t contain sharp boundaries. The use of directionality and window size in edge detection improves the boundary extraction result and thus directional Sobel 5x5 edge detection. This modification is the basic input for object based analysis because without appropriate segmentation they will not be able to extract the land objects. The use of feature selection by merging the qualitative and quantitative knowledge defines the rule sets thus the object based analysis shows that automatic land information extraction is possible with preservation of boundary. One can simply use the distinguishable properties to write the algorithm for each land class.

Arun and Katiyar (2013) presented an object extraction methodology in which cellular neural network was used for modelling feature shape, and adaptive kernel strategy along with corset optimization they are adopted to make the technique unsupervised. Salient features of this work are cellular automata approach based on adaptive kernel strategy intelligent compared with contemporary approaches using satellite images of Bhopal and Chandrapur cities in India.

They proposed a frame work for accurate detection of features along with automatic interpolation, and interpretation by modelling feature shape as well as contextual knowledge using advanced techniques such as SVRF, Cellular Neural Network, Corset, and MACA. The developed methodology was compared with contemporary methods using different statistical measures. Investigations over various satellite images revealed that considerable success was achieved with the CNN approach. CNN has been effective in modelling different complex features effectively and complexity of the approach has been considerably reduced using corset optimization. The system has dynamically used spectral and spatial information.

They observed that investigations of feature extraction process over various satellite images revealed that considerable success can be achieved with CNN approach. And that the approach can as well accurately detects various features better than existing methodologies and accuracy was revealed by extraction of air strip, roads and rivers. Materials used include: Google earth, Differential Global Positioning System (DGPS) and ERDAS. The visual results of few features extracted by system are given below, which also reveals Features extracted from LISS 3 sensor image, road network extracted from PAN sensor image and River extracted from LANDSAT sensor image.

Gap: The result shows that CNN based approach could be effectively model feature shapes and context sensitivity. Investigations have revealed that the method outperforms CA based feature shape modelling. Complexity of the approach has been considerably reduced using corset based approximation. Proposed system has proved to be intelligent with reference to accurate interpolation and interpretation. Disambiguation of features, enhanced detection, self-learning, minimal human interpretation, and reliability features of the system.

Ohlhof, Gulch, Muller, Wiedemann and Torre, (2015) conducted a research on the “Semi-Automatic Extraction of Line and Area Features from Aerial and Satellite Images”.These researchers carried out their project with two software companies ESG and Inpho. They developed an operational system for the semi-automatic extraction of line and area features in 2D and 3D based on an existing software platform. The complete system was delivered to the Geo-Information Office (AGeoBw) of the German Federal Armed Forces and has been in practical use since May 2003 for the update of VMap Level 1 data and the generation of the military basic vector database.The extraction tools for line and area features are integrated into Inpho’s software platform inJECT, which was originally designed for the measurement of 3D building models in digital imagery. For the import and export of the GIS vector data an interface between inJECT and the GIS packages Dynamo and GeoMedia (Intergraph), they are developed based on the GML2 format standard.

The semi-automatic extraction was preferably done in digital orthophotos for the capture of 2D GIS vector data. In addition, the software is available for the capture of 3D features using oriented aerial imagery. The algorithms and workflows have been extensively tested with IKONOS 2 and IRS satellite imagery as they will as orthophotos with 50 cm pixel size.

Gap: Inject is one of the very few software developments in automated feature extraction that have been implemented as a commercial system. The software has been substantially extended with new automation modules and a strong GIS interface. The software basis of inJECT has been proven to be an excellent platform to add and test external software modules that increase the automation level significantly. By adapting the OGC-defined GML standards the field for future applications is wide spread. The developed GIS interface to Dynamo and Geo-Media opens the window to many applications in those fields. Their solution has major advantages compared to many GIS data acquisition packages that do not allow such user-friendly data capture and updating, especially not in 3D. It has been demonstrated that existing vector data can be easily imported, updated and stored as well as attribute values added in the requested ways. The basic 3D building extraction kernelin inJECT has been extended by adding feature attributes to the geometrical features. The possibility to measure roads and parcels with a high level of automation has substantially increased the applicability for a wide range of users. The automated 2D extraction modules used for that purpose have been tested with several types of satellite imagery with a ground pixel size of 0.8 to 5m as well as aerial orthophotos of about 0.5 m ground resolution. But the results have not been compared with other existing Software.

Do, Pham, and Do, (2005) created a new algorithm that is very fast for building incremental, parallel and distributed SVM classifiers. Support Vector Machines (SVM) and kernel related methods have shown to build accurate models but the learning task usually needs a quadratic programming, so they observed the learning task for large datasets requires big memory capacity and a long time. A new incremental, parallel and distributed SVM algorithm using linear or nonlinear kernels proposed in this paper aims at classifying very large datasets on standard personal computers. They extend the recent finite Newton classifier for building an incremental, parallel and distributed SVM algorithm. The new algorithm is very fast and can handle very large datasets in linear or nonlinear classification tasks. An example of the effectiveness is given with the linear classification into two classes of two million datapoints in 20-dimensional input space in some seconds on ten personal computers (3 GHz Pentium IV, 512 MB RAM, Linux). The accuracy of the parallel and distributed incremental finite stepless Newton SVM algorithm is exactly the same as the original one.

To evaluate the performance of their parallel and distributed incremental finite stepless Newton SVM algorithm on very large datasets, they have implemented it on the PC running Linux Fedora Core 3. The program toolkit is written in C/C++. They have also used the high performance linear algebra library, Lapack++ to benefit by the high speed of computational matrix (Dongara, Pezo and Walker, 1993). Thus, the software program was able to deal with large datasets in linear and nonlinear classification tasks. They focus on numerical tests with large datasets generated by the RingNorm program (Delve, 1996).

By varying the number of PCs, the size of datasets and the number of dimensions, they have measured computational time. Thus, the parallel and distributed incremental algorithm has linear dependences on the number of machines, size of datasets and a second order of the number of dimensions. Concerning the communication cost, they take about one second when the dataset dimension is less than 100. The algorithm has linearly classified one million datapoints with 20-dimensional input space on ten machines in 2.585. The results obtained presented the effectiveness of these new algorithms to deal with very large datasets on PCs with linear and nonlinear classification tasks.

Gap: The result presented a new SVM algorithm being able to deal with very large datasets in linear and nonlinear classification tasks on PCs. They have extended the recent finite stepless Newton SVM algorithm to build incremental, parallel and distributed SVM. The accuracy of the new algorithm is exactly the same as the original one but its complexity is linearly dependence on the number of machines, size of datasets and a second order of the number of dimensions.The algorithm also requires storing a (n+1) x (n+1) matrix and two (n+1) x 1 vector in memory (where n is the number of dimensions). They focused on numerical tests with large datasets generated by the RingNorm program. Their new algorithm was also implemented by XMLRPC Kidd, 2001 which can do the parallel and distributed operations over any XML-capable transport protocol, typically over HTTP. The algorithm developed need to be tested using real-world satellite image dataset.

Iovan, Boldo and Cord (2008) presented a complete image analysis system, from high-resolution colour infrared (CIR) digital images, and a Digital Surface Model (DSM) so that

Detection, Characterization and Modelling Vegetation in Urban Areas from High Resolution Aerial Imagery can be carried out.The process starts with the extraction of all vegetation areas using a supervised classification system based on a Support Vector Machines (SVM) classifier. The result of this first step was further use to separate trees from lawns using texture criteria computed on the DSM. Tree crown borders are identified through a robust region growing algorithm based on tree-shape criteria.

A SVM classifier gives the species class for each tree-region previously identified. This classification was used to enhance the appearance of 3D city models by a realistic representation of vegetation according to the vegetation land use, shape and tree species. They compared the results obtained by applying state-of-the-art techniques for vegetation detection. They presented a complete hierarchical image analysis system to characterize urban vegetation which works on colour infra-red aerial images and contains components dealing with vegetation in urban areas, from extraction to tree species classification, thereafter, lawns were separated from trees, then tree crown borders were delineated and trees were classified according to their species. The proposed approach operates on standard aerial data and performs a complete characterization of vegetation in urban areas without any supplementary source of information, such as hyper-spectral or LiDAR (Light Detection and Ranging) data.

Gap: Research in the field of urban remote sensing often lacks tree species information. Their study describes a novel application of image texture analysis to classify tree crowns according to their species. The results were promising, pointing towards future large-scale classification of vegetation in human settlements from high-resolution aerial images.

Liu, Chang and Ma (2014) discussed the application of SVM in water quality assessment and uses SVM method to assess karst groundwater quality at the Niangziguan fountain region of Haihe River basin. Eight assessment factors were selected to randomly generate sample set. All test samples were classified correctly after training the model. The result shows that such a method solves the complex nonlinear relationship between assessment factor and water quality grade. It offers high prediction accuracy and is a reasonable and feasible assessment method.

Since water quality is influenced by many factors, these methods do not solve the complex nonlinear relationship between assessment factor and quality grade, and the assessment result is largely affected by the subjective factors of the assessing person. Applying SVM in water quality assessment, the multiple-factor water quality assessment model based on SVM was established. Their result shows that during the last 10 years, karst water has been increasingly contaminated and water quality dropped, mostly revealed by the rising level of total hardness and sulphate. Nitrate and nitrite level fluctuate with the precipitation without significant increasing in general.

Gap: Their research establishes SVM water quality assessment model. The model maps actual issues onto high-dimensional feature space using nonlinear conversion. It has good recognition capability on the characteristics of samples. The calculation result of the model shows that SVM has favourable classification performance and can be applied in water quality assessment.

Decherchi, Ridella and Zunino, (2009), developed a general method to compute the generalization bounds for SVMs, which is based on referring the SVM parameters to an unsupervised solution, and shows that such an approach yields tight bounds and attains effective model selection. When one estimates the generalization error, one uses an unsupervised reference to constrain the complexity of the learning machine, thereby possibly decreasing sharply the number of admissible hypothesis.They adopted Vector Quantization (VQ) as a representation paradigm, and introduce a biased regularization approach in bound computation and learning. Experimental results validate the proposed method on complex real-world data sets.

To overcome the hindrance of maximum-discrepancy (MD), they proposed a criterion to identify and sort a subset of admissible functions within the considered general model. The method uses unsupervised learning to derive a ‘reference’ classifier, and the SVM parameters trained on the actual targets to set a limiting boundary. Only the SVM classifiers that lie within that boundary, with respect to the reference, are admissible when computing the generalization bound and also described an express procedure for implementing the optimization process under the constrained-capacity mechanism. The effectiveness of the method was illustrated by using real world datasets; results confirmed that the constraint based approach leads to optimal hyper-parameters and featured tighter bounds than the conventional SVM method.

Gap: They constrained SVM by using VQ results as a reference and the methodology shows a wider general validity, and the approach can be extended to other classifier models. Since choosing a certain reference solution is equivalent to choosing a ‘prior’, other reference solutions than VQ can be adopted. They initiated the refinement of SVM, by using biased regularization for the computation of generalization bounds.

Huysmans, Baesens, Vanthienenl and Gestel (2006) have a projection of SVM predictions onto a Self-Organizing Maps (SOM) to gain more insight in the SVM model and the use of multi-year data sets to study evolutions in the perceived level of corruption. To achieve this, data from three different sources was combined. Demographic information, for example literacy and infant mortality rate, was retrieved from the CIA Factbook (http://www.cia.gov/cia/publications/factbook/) together with macro-economic variables, like GDP per capita and sectorial GDP information.

Information concerning the corruption level in specific countries was derived from Transparency International (http://www.transparency.org/) under the form of the Corruption Perceptions Index (CPI). This index ranks countries according to the degree to which corruption is perceived to exist among public officials and politicians. The CPI is created by interviewing business people and country analysts and gives a score between 0 (highly corrupt) and 10 (highly clean) to each country. In this study, data concerning the years 1996, 2000 and 2004 was used. In the index of 1996, 54 countries received a corruption score. They select only these countries and omit from the more elaborated 2000 and 2004 indices all other countries, resulting in a total of 162 observations: three observations from different years for each of the 54 countries.

In their SOM Analysis, they trained a self-organizing map of 15 by 15 neurons and all the available variables, including the CPI-scores; their results have shown that SOMs are a suitable tool for the exploration of data sets. They performed the traditional k-means clustering algorithm on the trained map of 15 by 15 neurons. The result of this procedure, assuming 5 clusters are present in the data. They integrated SOM and SVM and observed that the LS-SVM model provides a better forecasting accuracy than the corresponding linear model. However, they noted that the LS-SVM has a serious disadvantage: it is very difficult to understand the motivation behind this model’s decisions due to its complexity. To relieve this opacity restriction and to gain more insight in the model’s behaviour they projected its forecasts and the forecasting errors onto a self-organizing map.

In the end, a data mining approach for the analysis of corruption was presented. Where, the powerful visualization possibilities of self-organizing maps were used to study the interconnections between various macro-economic variables and the perceived level of corruption. The use of multi-year data sets allowed them to visualize the evolution of corruption over time for specific countries. And secondly they reported that forecasting models can be constructed that allow analysts to predict the level of corruption for countries where this information is missing.

Gap: Most of the existing literature discusses the topic from a socio-economical perspective and only few studies tackle this research field from a data mining point of view. They applied data mining techniques onto a cross-country database linking macro-economic variables to perceived levels of corruption. In the first part, self-organizing maps was applied to study the interconnections between these variables. Afterwards, support vector machines are trained on part of the data and used to forecast corruption for other countries. Large deviations for specific countries between these models' predictions and the actual values can prove useful for further research. Also, there was a projection of the forecasts onto a self-organizing map which allows a detailed comparison between the different models' behaviour.This ideology should be evaluated using remote sensing data and applied in various remote sensing real-world applications.

L'uborLadický, Torre and Zizerman (2012), investigated the Latent SVMs for Human Detection with a Locally Affine Deformation, in doing these, they proposed a latent deformable template model with a locally affine deformation field, which allows for more general and more natural deformations of the template while not over-fitting the data; and also provide a novel inference method for this kind of problem. The deformation model was to measure the distances between training samples, and their result show how this can be used to cluster the problem into several modes, corresponding to different types of objects, viewpoints or poses.

Their method leads to a significant improvement over the state-of-the-art with small computational overhead. The training samples are clustered into 10 models. Clustering of the training samples into multiple models together with their corresponding HOG template deformed by the latent deformation field was tested and positive detections are overlaid with the learnt HOG template of the corresponding model, deformed by the deformation field. Large majority of the persons are correctly detected. Typical mistakes include missed detections of hard instances (A3, A4, C3, and E4), false positive detections (A5, B2), detection with insufficient overlap with the ground truth (B3) or detection by a wrong model with a wrong truncation (A3).

Gap: the tested algorithm on the challenging Buffy data set showed promising results. They assume, their random field formulation with the locally affine constraints could be used in the future for other computer vision tasks.The deformation model developed can be test in building deformation monitoring and modeling.

Aymerich, Pierra, Soria-Frisch and Obermayer, (2010) presented a comparison between two different techniques for fast phytoplankton discrimination: Self-Organizing Maps (SOM) and Potential Support Vector Machines (P-SVM), evaluating its capability to achieve phytoplankton classification from its fluorescence spectra. Five different cultures representing the major algae divisions were selected and grown under the same conditions. Excitation-Emission Matrices were acquired every day.

Herein they present the results using P-SVM, and they compare the results using both techniques. Two data sets were selected: excitation spectra at a 680nm emission wavelength, and emission spectra from samples excited at 470nm. In the first case, using excitation spectra, training and test data sets were chosen doing repeated random sub-sampling validation, and the results of 10 different classifications were averaged. In order to evaluate the performance of techniques, the confusion matrices from the classification step were obtained, and the index Kappa by Aymerich, Piera and Soria-Frisch undated. was computed. The averaged Kappa index resulted is 0.3636. SOM performed better in this case obtaining K=0.6629.

Using emission fluorescence spectra instead of excitation spectra, training and test data sets were chosen by random sub-sampling validation, and the results of 10 different classifications were averaged. The results in that case increase slightly, obtaining K=0.4839, while using SOM they obtained 0.6568.Once the data were pre-processed, P-SVM was used again to classify the different spectra into the correct class. The index Kappa obtained this time was 0.7141. In contrast to the results obtained with SOM (K=0.6992), the performance of P-SVM using pre-processed data are even better than those obtained with SOM.

Gap: Utilizing SOM shows poor results and could be due to the similarity of the emission fluorescence spectra among the different classes studied. In order to enhance the fluorescence spectra differences between the algae, a derivative analysis has been applied to the emission spectra, where previous work shows that the derivative analysis has demonstrated to be a powerful tool to enhance differences (Butler, and D. W. Hopkins, 1970), although it is high sensitive to noise. For this reason, the noise of the spectra has been also reduced using a wavelet denoising technique according to Piera et al, 2004.

Song, Hu and Xie (2002)confirmed thatclassification of bullet-hole images is possible. They tested both the SVM algorithms using different kernel functions and regularization parameters. A classical region growth algorithm was used to obtain each independent bullet-hole image. Normally, certain bullet-hole images contain two or more bullet holes with different degrees of overlapping. Therefore, it is critical to classify bullet-hole images into multi-class sets for the auto-scoring system. The main task of the auto-scoring system was to classify the bullet-hole images into proper classes.

They also propose a robust support vector machine for pattern classification, which aims at solving the over-fitting problem when outliers exist in the training data set. The incorporation of the average techniques to the standard support vector machine (SVM) training makes the decision function less detected by outliers, and controls the amount of regularization automatically. Experiments for the bullet hole classification problem showed that the number of the support vectors was reduced, and the generalization performance was improved significantly compared to that of the standard SVM training.The larger the parameter is, the nearer the support vectors will be to those data points toward the centre. The algorithm becomes "more robust" against outliers and thus, results in smoother decision surface. However, the classification error may be increased correspondingly. The characteristic of this algorithm is to gain robustness against outlier at cost of the classification accuracy. A good selection of parameter is the one that leads to compromise between the robustness and classification accuracy.

Gap: They proposed a general robust SVM algorithm against outliers by adding the distance between each data point and the centre of classes to form the margin of separating hyperplane, good robust performance was achieved. The simulation of robust SVM with different kernel functions and regularization parameters was presented to show that the robust algorithm can be used for pattern classification problems with different difficulty levels. The experiment results show that the decision boundary becomes less dithered and the number of support vectors of the robust SVM were reduced significantly compared to that of the standard SVM.

Huang, (2011) noted that Support Vector Machine (SVM) has been widely used in the study of stock investment related topics and thus carried a Genetic Algorithm Optimization SVM to

Construction of Investment Model. According to the study, stock investment can be further divided into three strategies such as: buy, sell and hold and using data concerning China Steel Corporation, genetic algorithm for the search of the best SVM parameter and the selection of the best SVM prediction variable was adopted, and was compared with Logistic Regression for the classification prediction capability of stock investment.The researcher carried out this application using Matlab SVM function and genetic algorithm tool box.

From the classification prediction result it can be seen that the SVM after adjustment of input variables and parameters have classification prediction capability relatively superior to that of the other three models. Their result shows that a combination of genetic algorithm and SVM model can enhance the decision-making capability of the model, and it is thus very suitable for constructing stock investment model. However, general SVM model and Logistic Regression model, shows it is difficult to judge the superiority of these two models which means, if the input variable or parameter of general SVM model is not appropriately selected, it will not necessarily perform better than general statistical regression model.

Gap: The major contribution of their work was to address the appropriateness of model parameter and prediction variable can usually affect the prediction capability of the entire model. Genetic algorithm was applied to adjust model parameter and select prediction variable in this research. In addition, classification capability comparison and analysis of stock investment was done through SVM model and Logistic Regression model.

Anuja-Kumari and Chitra (2013) developedSVM models for the classification of diabetes dataset. The experiments were conducted on Matlab R2010a. The datasets were stored in MS Excel documents and read directly from Matlab. Pima Indian diabetes dataset, donated by Vincent Sigillito, was a collection of medical diagnostic reports from 768 records of female patients at least 21 years old of Pima Indian heritage, a population living near Phoenix, Arizona, USA formed the data base for the analysis. The binary target variable takes the values „0" or „1" while „1" means a positive test for diabetes, „0" means a negative test. There are 268 cases in class „1" and 500 cases in class „0". The significance of the automatically selected set of variables was further manually evaluated by fine tuning parameters. The variables included in the final selection are those with the best discriminative performance.

To evaluate the robustness of the SVM models, a 10-fold cross-validation was performed in the training data set. The experimental results obtained show that support vector machine can be successfully used for diagnosing diabetes disease.This cross-validation process was repeated 10 times, and each subset serves once as the test data set. Test data sets assess the performance of the models.The Pima Indian diabetic database at the UCI machine learning laboratory has become a standard for testing data mining algorithms to see their prediction accuracy in diabetes data classification.

Gap: They used datasets for diabetes disease from the machine learning laboratory at University of California, Irvine and trained the data by using SVM. The choice of best value of parameters for particular kernel is critical for a given amount of data. SVM approach can be successfully used to detect a common disease with simple clinical measurements, without laboratory tests. In the proposed work, SVM with Radial basis function kernel is used for classification. The performance parameters such as the classification accuracy, sensitivity, and specificity of the SVM and RBF have found to be high thus making it a good option for the classification process. In future the performance of SVM classifier can be improved by feature subset selection process.

Anis and Abdelaziz (2012) studied thecomplex support vector machine regression for robust channel estimation inLTE downlink system.This study depicts the problem of channel estimation for LTE Downlink system in the environment of high mobility; presenting non-Gaussian impulse noise interfering with reference signals is faced. The estimation of the frequency selective time varying multipath fading channel was performed by using a channel estimator based on a nonlinear complex Support Vector Machine Regression (SVR) which is applied to Long Term Evolution (LTE) downlink. The estimation algorithm makes use of the pilot signals to estimate the total frequency response of the highly selective fading multipath channel. Thus, the algorithm maps trained data into a high dimensional feature space and uses the structural risk minimization principle to carry out the regression estimation for the frequency response function of the fading channel. The obtained results show the effectiveness of the proposed method which has better performance than the conventional Least Squares (LS) and Decision Feedback methods to track the variations of the fading multipath channel.

Gap: They describes a nonlinear complex SVR based channel estimation technique for a downlink LTE system in the presence of Gaussian noise and impulse noise interfering with OFDM pilot symbols in high mobility environment. The proposed method is based on learning process that uses training sequence to estimate the channel variations. Their formulation is based on nonlinear complex SVR specifically developed for reference-based

OFDMA systems. Simulations have confirmed the capabilities of the proposed nonlinear complex OFDM-SVR estimator in the presence of Gaussian and non-Gaussian impulse noise interfering with the reference symbols for a high mobile speed when compared to conventional LS and Decision feedback methods for a highly selective time varying multipath fading channel.The algorithm developed need to be tested using real-world satellite image dataset for change detection modeling or in enhancing the global positioning system (GPS) signal.

Yu, Yang and Han, (2003)aimed at ‘’Classifying Large Data Sets Using SVMs with Hierarchical Cluster’’ On this they proposed a new method called CB-SVM (Clustering-Based SVM) that integrates a scalable clustering method with an SVM method and effectively runs SVMs for very large data sets. CB-SVM applies a hierarchical micro-clustering algorithm that scans the entire data set only once to provide an SVM with high quality samples that carry the statistical summaries of the data such that the summaries maximize the benefit of learning the SVM. CB-SVM tries to generate the best SVM boundary for very large data sets given limited amount of resources. Their experiments on synthetic and real data sets show that CB-SVM is highly scalable for very large data sets while also generating high classification accuracy.

They noted that the existing SVMs are not feasible to run such data sets due to their high complexity on the data size or frequent accesses on the large data sets causing expensive I/O operations. CB-SVM applies a hierarchical micro-clustering algorithm that scans the entire data set only once to provide an SVM with high quality micro-clusters that carry the statistical summaries of the data such that the summaries maximize the benefit of learning the SVM. CB-SVM tries to generate the best SVM boundary for very large data sets given limited amount of resource based on the philosophy of hierarchical clustering where progressive deepening can be conducted when needed to find high quality boundaries for SVM. Their experiments on synthetic and real data sets show that CB-SVM is very scalable for very large data sets while generating high classification accuracy.

Gap: They observed that other researchers have proposed various revisions of SVMs to increase the training efficiency by mutating or approximating it. However, they are still not feasible with very large data sets where even multiple scans of the entire data set are too expensive to perform, or they end up losing the benefits of using an SVM by over-simplifications.

Jin, Feng and Shen, (2011), carried an automatic road feature extraction approach and applied it to a number of aerial images in urban areas. Their aim was to produce an accurate urban road model reconstruction from high resolution remotely sensed imagery based on Support Vector Machine and Gabor filters, and so supervised SVM image classification technique was employed to segment the road surface from other ground details, and the road pavement markings are detected on the generated road surface with Gabor filters. Their experiments using several pan-sharpened aerial images of Brisbane Queensland, has validated the proposed method were the objective of the experiment was to determine the performance of the proposed road feature extraction approach quantitatively over the study area.The training data used to train the support vector machine and the resulting model was used to classify the whole image into two features: road and non-road. For their implementation of SVM, the software package LIBSVM by Chang and Lin (2001) was adapted. Gaussian RBF was used as the kernel function, and the parameter C was set to be 10. After the image classification, the connected component analysis was used to remove small noises misclassified into road class.Finally, the Gabor filtered image is then segmented by Otsu’s thresholding algorithm, and directional morphological opening and closing algorithms are utilized to remove misclassified features. The quantitative evaluation of the experimental results was carried out by comparing the automated (derived) results against a manually compiled, high quality reference model.

The road marking accuracy evaluation was carried out by comparing the extracted pavement markings with manually plotted road markings used as reference data as presented in Wiedemann et al., (1998), and both data sets were given in vector representation. The buffer width is predefined to be the average width of the road markings, and they set it to be 15 cm in their experiment. Road boundaries and road markings are firstly digitized from the test images and used as ground truth. For their entire four test sites, nearly 90% of the road surfaces are correctly detected, and the relevant false alarm rate is about 10%. The completeness of road pavement marking extraction reaches above 87%, except for test site IV, which is seriously affected by shadows. The shadows on the road surfaces can reduce the intensity contrast between pavement markings and the road surfaces background, which makes it difficult to enhance the road markings by Gabor filter. The average false alarm rate of the four test sites is about 10%.

Gap: An automatic road surface and pavement marking extraction approach from aerial images with high spatial resolution was proposed. The developed method was based on SVM image classification as well as Gabor filtering, and can generate accurate lane level digital road maps automatically. The experimental results using the aerial image dataset with ground resolution of 0.07 m have demonstrated that the proposed method works satisfactorily.

Yao, Yang, Yongqing Hong WeimingLv and Chen, (2013) used a novel approach based on clustering algorithm, in which only a small subset was selected from the original training set to act as their final training set. Their algorithm works to select the most informative samples using K-means clustering algorithm, and the SVM classifier is built through training on those selected samples. Experiments show that their approach greatly reduces the scale of training set, thus effectively saves the training and predicting time of SVM, and at the same time guarantees the generalization performance.

In their experiments, they put efforts on solving the binary classification problem, and the training dataset is collected from the original dataset through stratified sampling. Their developed algorithm and LIBSVM are implemented on three datasets, respectively. Each dataset is divided into two parts, the training set and the testing set. And the experiment results show that the number of SVs decrease greatly and the predict accuracy is the same or a litter higher than that of LIBSVM. The algorithm was implemented with Matlab and libsvm toolbox. The kernel function they used here is the Gaussian kernel

Gap: They were dedicated to solve the problem of reducing the scale of the training set with the method of clustering. They used the K-means clustering approach to select the few most informative samples, which are used to construct their real training set. Experiment results show that their algorithm reaches the goal of reducing the scale of training set, and greatly reduces the training and predicting time, meanwhile assures the generalization ability of their K-SVM algorithm.

Radha and Rehna (2014) carried out a research on the Retrieval of Geographic Images Using Multi Support Vector Machine and Local Invariant Features**.** They proposed an approach was based on Multi-SVM algorithm. It makes use of Scale Invariant Feature Transform (SIFT). First the input image or target image that has to be detected is identified. The next step is to detect the key-points in the image. After detecting the key-points then feature extraction is done. It was done by using simple statistics, homogenous texture and colour histogram. Then by using ‘Euclidean Distance Measure’ the most similar images were retrieved. Multi-SVM algorithm was used to perform image retrieval and using this algorithm the false image retrieval was considerably reduced. An SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. Local features allow a greater range of objects and spatial patterns to be observed. SIFT detector was used to identify the key-points in the image. Using the proposed algorithm, the image retrieval gives accurate results.

Gap: They presented an investigation into local invariant features for overhead image retrieval, the first such study of its kind. They demonstrated that local invariant features are more effective than standard features such as colour and texture for image retrieval of LULC classes in high-resolution aerial imagery. They performed image retrieval using Multi Support Vector Machine and it was seen that the false image retrieval has been considerably reduced. However, how these results were validated was not stated.

Platt (1998) proposedSequential Minimal Optimization (SMO) as a fast algorithm for training support vector machines. The researcher noted that training a support vector machine requires the solution of a very large quadratic programming (QP) optimization problem and SMO can breaks this large QP problem into a series of smallest possible QP problems. These small QP problems are solved analytically, which avoids using a time-consuming numerical QP optimization as an inner loop. The amount of memory required for SMO was linear in the training set size, which allowed SMO to handle very large training sets. Because matrix computation was avoided, SMO scales somewhere between linear and quadratic in the training set size for various test problems, while the standard chunking SVM algorithm scales somewhere between linear and cubic in the training set size. SMO's computation time is dominated by SVM evaluation; hence SMO is fastest for linear SVMs and sparse data sets. On real-world sparse data sets, SMO can be more than 1000 times faster than the chunking algorithm.

The SMO algorithm is then presented in detail, including the solution to the analytiheuristics for choosing which variables to optimize in the inner loop, a description of how to set the threshold of the SVM, some optimizations for special cases, the pseudo-code of the algorithm, and the relationship of SMO to other algorithms.SMO has been tested on two real-world data sets and two artificial data sets. Thus, they present the results for timing SMO versus the standard "chunking" algorithm for these data sets and presents conclusions based on these timings.

Gap: For the various test sets, the training time of SMO empirically scales between ~N and ~N 2.2. The training time of chunking scales between ~N 1.2 and ~N 3.4. The scaling of

SMO can be more than one order better than chunking. For the real-world test sets, SMO can be a factor of 1200 times faster for linear SVMs and a factor of 15 times faster for non-linear SVMs.Because of its ease of use and better scaling with training set size, SMO is a strong candidate for becoming the standard SVM training algorithm. More benchmarking experiments against other QP techniques and the best Osuna heuristics are needed before final conclusions can be drawn.

Psaltis, and Ioannidis, (2010) designed amethodology which illustrated how to accomplish supervised change detection on simulated data employing support vector machines. The main idea of the proposed procedure follows the supervised classification paradigm. The first step is to layer the available data for the same region in different time periods. Then evaluate a number of predefined cues for the whole region and use some manually collected positive and negative samples to train a classifier. Finally, this classifier can be used to assert change in the remaining data. The base data chosen were very high resolution orthoimages and digital surface models (DSMs) because they offer both the radiometric and geometric information needed for robust change detection. In an effort to improve the procedures for automated and reliable detection of new buildings, which can be used for several applications such as map updating, monitoring of informal development, a procedure using Support Vector Machines was developed and tested. The results using the Radical Basis Function SVM classifier are especially encouraging, where simple noise conditions have been applied, the results were excellent. All new buildings were detected and few false replies were mentioned. Consequently, the proposed procedure works perfectly under ideal conditions. When the inserted noise becomes complicated (two step noise) the success indicators are reduced (e.g., detection of new buildings 80% and 5% false negative responses). The use of DSMs seems to be necessary; DSMs of high quality and accuracy give the best results. When training and testing data differ in quality, the difficulties in detection of changes grow. It is proven that the use of images alone is no longer adequate (detection of new buildings is reduced to less than 70%). The use of DSMs is necessary; the increased quality of DSM leads to better results.

Gap: From these scenarios it was evident that the DSM and high DSM show positive impact to quality of the final result. As it was very crucial to train and test the SVM in almost the same conditions in order to achieve uniform levels of acceptable accuracy. If this is not possible then a high quality DSM can greatly enhance the overall performance of the algorithm.

Boley and Caoy (2004) whoaimed at training support vector machine using adaptive clustering,noted that the training of support vector machines involves a huge optimization problem. And thus, proposed another algorithm called Cluster SVM that accelerates the training process by exploiting the distributional properties of the training data, that is, the natural clustering of the training data and the overall layout of these clusters relative to the decision boundary of support vector machines. The proposed algorithm first partitions the training data into several pair-wise disjoint clusters. Then, the representatives of these clusters were used to train an initial support vector machine, based on which they can approximately identify the support vectors and non-support vectors. After replacing the cluster containing only non-support vectors with its representative, the number of training data can be significantly reduced, thereby speeding up the training process. The proposed Cluster SVM has been tested against the popular training algorithm SMO on both the artificial data and the real data, and a significant speedup was observed. The complexity of Cluster SVM scales with the square of the number of support vectors and with improvement, it is expected that it will scale with square of the number of non-boundary support vectors.

Gap: The experiment shows that the training time of this strategy scales with the square of the number of support vectors and an approximate solution can be found even faster. Further, based on the theory underlying Cluster SVM, it is expected that the training time will scale with the number of boundary support vectors after some straightforward extensions to the current work.

Suykens, Brabanter, Lukas and Vandewalle, (2001) on a project called “Weighted Least Squares Support Vector Machines: Robustness and Sparse Approximation”,however,observed thatLeast squares support vector machines (LS-SVM) is an SVM version which involves equality instead of inequality constraints and works with a least squares cost function. In this way, the solution follows from a linear Karush–Kuhn–Tucker system instead of a quadratic programming problem. But, sparseness is lost in the LS-SVM case and the estimation of the support values is only optimal in the case of a Gaussian distribution of the error variables. They discussed a method which can overcome these two drawbacks. Their methods illustrated for RBF kernels and demonstrate how to obtain robust estimates with selection of an appropriate number of hidden units, in the case of outliers or non-Gaussian error distributions with heavy tails.

Their work shows that one can overcome these drawbacks concerning sparseness and robustness within the present LS-SVM framework and how one can apply weighted least squares in order to produce a more robust estimate. This was done by first applying an un-weighted LS-SVM and, in the second stage, associate weighting values to the error variables based upon the resulting error variables from the first stage. Techniques of weighted least squares are well known e.g. in statistics, identification and control theory and signal processing. Furthermore, they illustrated how sparseness can be imposed to the weighted LS-SVM solution by gradually pruning the sorted support value spectrum.

Gap: They have shown how to obtain robust estimates within the LS-SVM framework in the case of outliers and heavy tailed non-Gaussian error distributions. This was done by applying a weighted LS-SVM version. While least squares in standard parametric linear regression has a low breakdown point, LS-SVM’s with RBF kernel have much better properties. Nevertheless, the robustness can be further enhanced by applying an additional weighted LS-SVM step. While standard SVM’s possess a sparseness property for the solution vector, this feature is lost when considering LS-SVM’s. However, they have shown how sparseness can be obtained by pruning the sorted support value spectrum if needed. The pruning procedure is based upon the solution vector itself which is an advantage in comparison with classical multilayer perceptron pruning. This procedure has the potential advantage of keeping the hyper parameter selection more localized. While standard SVM approaches start from choosing a given convex cost function and obtain a robust and sparse estimate in a top–down fashion, this procedure has the disadvantage that one should know in fact beforehand which cost function is statistically optimal. They have successfully demonstrated an alternative bottom–up procedure which starts from an un-weighted LS-SVM and then makes the solution robust by defining weightings based upon the error distribution. This provides motivation for further fundamental research in this direction for future work.

Mahi, Hadria and Chahira, (2014) Zernike moments-based descriptor was used as a measure of shape information for the detection of buildings from Very High Spatial Resolution (VHSR) satellite images. Their proposed approach comprises three steps. First, the image was segmented into homogeneous objects based on the spectral and spatial information. Mean Shift segmentation method was used for this end. Second, a Zernike feature vector was computed for each segment. Finally, a Support Vector Machines (SVM)-based classification using the feature vectors as inputs was performed. Experimental results and comparison with Environment for Visualizing Images (ENVI) commercial package confirm the effectiveness of the proposed approach.

The first experiment was conducted using a synthetic gray-scale image with sixteen distinct regions. In order to classify this image using only shape information, the first step was to extract homogeneous segments by applying the Mean Shift segmentation algorithm. After the image was segmented, a Zernike feature vector was computed for each of the segments in the image. Then, t vectors are used as inputs to the SVM classifier with an RBF kernel. Seven samples were selected according to the segment shape for training the SVM.

The second experiment was conducted using a subscene of Quickbird image datasets of Algiers (northern Algeria). This one was acquired on August consists of three multispectral images with a spatial resolution of 0.61m. The result shows the extracted buildings classes with the original image as background: buildings #1 (red colour), buildings #2 (green colour) and buildings #3 (blue colour). It can be seen that overall most of building segments were correctly identified with a good visual quality. Indeed, the Building segments are homogeneous and look like the reality. By visual inspection, the detection accuracy is more than 83% in the case of buildings #1 and is more than 84% in the case of buildings #2. Here the accuracy is defined as the ratio between number of buildings classified correctly and total buildings in the sub-scene.

However, the analysis of the results shows that the most important source of errors lies with confusion between buildings and grass areas (red and blue dashed regions). These confusions are mainly due to the fact that grass areas and buildings have similar shapes. This problem can be avoided by computing the Normalized Difference Vegetation Index (NDVI). Indeed, the regions whose average NDVI values are higher than an NDVI threshold are denoted as vegetation areas. In this way, vegetation areas can be eliminated. It can also be seen from the results that the class building #4 was misclassified. This drawback is mainly related to the segmentation result.

In fact, the object in the original image is formed by heterogeneous spectral regions and it is very difficult to find the best set of parameters values to obtain one segment for this object without loss of other best segments. To assess the performance of the proposed method, a comparative evaluation against the ENVI feature extraction tool was performed. The same training samples were used. For the case of the synthetic image, SVM classifier with RBF kernel was adopted.

Gap: A new approach for automated buildings extraction from VHRS images was proposed. First, image objects extraction was performed using the Mean Shift segmentation algorithm.

Then, a Zernike moments-based descriptor was calculated for each object or segment. Finally, a SVMs-based classification using the feature vectors as input instead of the original objects was carried out to assign a class label to each of the segments. The main outcome of this work was the use of Zernike moment's descriptor as a measure of shape information. Consequently, by performing the classification on the obtained Zernike feature vectors instead of the original objects, a good discrimination between object shapes can be achieved. The preliminary experimental results clearly demonstrated the potentials and the efficiency of the proposed approach as well as some application limitations, which encourages new reflections both for method enhancement and for future research.

Hsu and Lin (2002) incomparison of methods for multiclass support vector machines, observed thatthe Support vector machines (SVMs) were originally designed for binary classification and how to effectively extend it for multiclass classification is still an on-going research issue. Several methods have been proposed where typically they construct a multiclass classifier by combining several binary classifiers. Some authors also proposed methods that consider all classes at once. As it is computationally more expensive to solve multiclass problems, comparisons of these methods using large-scale problems have not been seriously conducted. So they used decomposition implementations for "all-together" methods. They compare their performance with three methods based on binary classifications: "one-against-all," "one-against-one," and directed acyclic graph SVM (DAGSVM). Their experiments indicate that the "one-against-one" and DAG methods were more suitable for practical use than the other methods. Results also show that for large problems methods by considering all data at once in general need further support vectors.

Gap: The experiments indicate that the "one-against-one" and DAG methods were more suitable for practical use than the other methods. Results also show that for large problems, methods of considering all data at once in general need further support vectors.

2.2.2 Review of studies done in Nigeria: local, regional and national

Ndehedehe (2014) applied the use of Support Vector Machine Based Kernel types in Extraction of Urban Areas in Uyo Metropolis from Remote Sensing Multispectral Image. The Metropolis lies within latitudes 40° 56' 30" N and 50° 07' 40" N, and longitudes 70° 49' 50" E and 80° 01' 00" E and is situated about 55 km inland from the coastal plain of South-Eastern Nigeria. Uyo Metropolis is the capital of Akwa Ibom State of Nigeria and is about 312.6 Sq. km with a population of about 400,000.Ndehedehe applied a pixel based classification to classify built up areas using a carefully-extracted training sample as a spectral signature of the specified region of interest. The SVM kernel type implementation was software-based and the results were validated using Orthophoto-derived vector of the study area. Then, the classified data using various SVM kernel types were vectorised and compared with the existing vector map of the same location using a GIS approach. An impervious surface map of the study area with very high accuracy was produced to quantify the exact pixels representing built environment from multispectral Landsat-based imagery.

Sub-pixel estimation of Impervious Surface Areas (ISA) was done by first using the high resolution data (Orthophoto) to calculate the proportional impervious cover for the specified region of interest. Regions of Interest (ROIs) were also used to extract statistics and average spectra from groups of pixels. The researcher created the ROIs of pixels and then examined and extracted statistics of the selected ROIs. Post-classification comparison was to be adopted to examine the performance of the SVM kernel types in the quantification of urban areas in Uyo metropolis.

Gap: This work looks into the performance of 4 SVM kernel types in the extraction of built up areas from multispectral remote sensing data. The classified urban segments of the various kernel types were compared with previously digitized vector of a section of the study area. The sigmoid kernel performed better than others in the quantification of built up areas. Nevertheless, the polynomial, linear and RBF kernels had a good agreement with the ground truth data. The SVM Sigmoid kernel trick has proved to be very useful in segmenting built environment   from multispectral satellite images.The researcher claims that the results were validated using Orthophoto-derived vector map of the study area, but however fail to test the accuracy of classification using the same software-based statistical tools like Confusion matrix and kappa's index.

Okwuashi, McConchie, Nwilo, Isong, Eyoh and Nwanekezie, (2012) carried out the **"**Predicting Future Land Use Change Using Support Vector Machine Based GIS Cellular Automata: A Case of Lagos, Nigeria". Because they observed that Lagos has undergone an unprecedented urban expansion and other contemporary findings favour the integration of cellular automata and geographic information systems for modelling land use change. Thus, the researchers introduced the support vector machine based GIS cellular automata calibration for land use change prediction of Lagos. The support vector machine based cellular automata model was loosely coupled with the geographic information systems. Support vector machine parameters were optimized with the k-fold cross-validation technique, using the linear, polynomial, and RBF kernels functions. The land use change prediction was based on three land use epochs: 1963-1978, 1978-1984, and 1984-2000. The performance of the model was evaluated using the Kappa statistic and receiver operating characteristic. The order of performance of the three kernels is: RBF, polynomial, and linear. The results indicate substantial agreement between the actual and predicted maps. The urban forms in 2015 and 2030 were predicted based on the three land use epochs.

The modelling was implemented in MATLAB and visualised in ArcGIS. The training data were extracted using the stratified random sampling. The training data consist of developed and undeveloped cells. Developed cells were labelled +1 while undeveloped cells were labelled -1. The historical land use change from 1963-1978, 1978-1984, and 1984-2000 was used to forecast the most probable land use maps in 2015 and 2030. The future land use maps in 2015 and 2030 were derived by running the SVM based CA model iteratively. The predicted maps in 2015 and 2030 using the polynomial, RBF, and linear kernels functions were depicted.

Gap: The computed ROC results corroborated with the kappa statistic results. The order of performance of the three kernel functions based on kappa and AUC estimates was: RBF, polynomial, and linear. The computed AUC and Kappa statistic results from this experiment indicate substantial agreement between the actual and the predicted maps of Lagos. The satisfactory results from this experiment imply that the support vector machine based GIS cellular automata model is a promising tool for predicting land use change.

Owamoyo, Alaba and Abimbola (2013) carried out “Number Plate Recognition for Nigerian Vehicles” using Sobel filter, morphological operations and connected component analysis. Character segmentation was done by using connected component and vertical projection analysis. Character recognition was carried out using Support Vector machine (SVM). The segmentation accuracy was 80% and recognition rate was 79.84 %. Their database consists of different sized JPEG coloured images which was resized to 1024 x 768 and total of 250 images were used to test the algorithm. The images were taken with different background as well as illumination conditions and experiments shows that the algorithm has good performance on number plate extraction, and character segmentation work. The images were analysed correctly, with noise, illumination variance, and rotation to ±50. This work was implemented using MATLAB.

However, the observed that deep shadows and reflections have an impact on number plate extraction work. Because of uneven illumination, stained number plates, true number plates could not get correctly extracted. Failure in character segmentation was mainly because of merging of characters on number plate, stained number plates, orientation of the image and poor illumination. Character recognition work was done on 10 digits (0 to 9) and 26 alphabets (A to Z). The recognition rate achieved using 'polynomial' kernel is 79.84 %. The false recognition is due to similarity in the character shape, e.g. 6 and B, 5 and S etc. Recognition can be increased if two groups of SVM's are designed separately, one for digits and other for alphabets.

A high resolution digital camera was used to acquire an image. Images are taken in different background, illumination conditions, and at various distances from the camera to vehicle. All the processing steps are executed on grayscale image. Pre-processing is mainly used to enhance the processing speed, improve the contrast of the image, and to reduce the noise in the image. In order to reduce the problem of low quality and low contrast in car images, images are enhanced by using histogram equalization on gray scale image.

Gap: An algorithm for vehicle number plate extraction, character segmentation and recognition was presented. Database of the image consists of images with different size, background, illumination, camera angle, and distance among others.The experimental results show that, number plates can be extracted faithfully based on vertical edge detection and connected component algorithm, with the success rate of 85%. Character segmentation phase using connected component analysis and vertical projection analysis works well with the success rate of 80%. The success rate achieved for character recognition is 79.84%. Recognition can be increased if two groups of SVM's are designed separately, one for digits and other for alphabets. The approach required to be tested using real-world remote sensing problem.

United Nations Human Settlements Programme (UN-HABITAT, 2009)in their work published as an Executive Summary of StructurePlans for Awka, Onitsha and Nnewiand Environs 2009-2027 they adopted the participatory and very engaging Rapid Urban Sector Profiling for Sustainability (RUSPS), first developed by UN-HABITAT team and has been successfully employed in several countries. The UN-HABITAT organized a week-long training on RUSPS to properly induct 10 Nigerian Experts and 15 governments technical staff drawn from Anambra State Ministries and local governments and Federal Ministry of Housing and Urban Development in RUSPS methodology. During the training, RUSPS global data collection framework was reviewed with an agreement on the under-mentioned seven thematic issues for the conduct of RUSPS for Anambra cities: Basic urban services; Environment, Gender, Governance, Heritage/historic areas, Local Economic Development; and Shelter and slums. The noted that Anambra, with its population of over 4 million people in 2006, is the second most urbanized states in the country, having 62% of its total population living in urban areas. Unfortunately, past Governments since creation of the State in 1991 have failed to adopt city development strategies for the many fast growing cities in the State to cope with rapid urbanization. Following decades of neglect and poor urban governance, the profiles of these cities indicate that they are characterized by decayed inner and suburban sprawling slums, inadequate sanitation, uncontrolled street trading, mountains of uncollected wastes, overcrowded and congested transport systems and roads with poor drainages, noise and air pollution. Their Nnewi Planning Area covered some 72 sq km including Ichi, Oraifite, Ozubulu all in Ekwusigo Local Government Area; Nnobi, Awka-Etiti in Idemili South Local Government Area; Amichi,Azigbo, Utu and Ukpor in Nnewi South LocalGovernment, with one-third of this area fully built.IKONOS Satellite image, Personal observation, statistical records and ArcGIS software was used in the analysis.

Gap: Image digitization and few statistical records were used but Image classification and validation was not carried out and questionnaires were not distributed to capture a larger view of the resident of these area or respondents of the study area.

2.2.3 Review of studies done outside Nigeria

Kwesi, (2012) mapped oil palm related land cover of a section from northern portion of Ejisu-Juaben district in the Ashanti Region of Ghana using support vector machine (SVM) with Landsat ETM+. The district lies within Longitude 6° 15ʺ N and 7° 00ʺ N and Latitude 1° 15ʺ W and 1 ° 45ʺ W and is characterised by both agricultural and socio-economic activities. The Landsat ETM+ data acquired in 2010 was used for processing and image classification. Field data were acquired in October 2011 through stratified random sampling. A total of 343 samples were collected for classification and accuracy assessment. The classification was carried out using MLC and SVM based on best three band combination from the image. The SVM and MLC performance evaluation was done using overall accuracy assessment and kappa statistics procedure. The results of separability analysis showed that ETM+ data provides spectral discrimination of land cover types found in the study area. The best three bands that provided the optimum spectral separability based on Bhattacharyya distance are 4, 5, and 3. The result showed that band 4, band 5 and band 3 provided best spectral separability. The overall accuracy result of the SVM classification was 78.29% (kappa statistic = 0.73). The RBF parameter setting in SVM was an important variable in the classification process, because it helped control the number of support vector used in the classification. The overall accuracy for MLC was 71.7% (kappa statistics =0.65).

Gap: The study was conducted in Bomfa, Apemso, Kote, Apraku, Juaben, and Ejisu farming communities with favourable agro-climatic conditions; located within the northern portion of Ejisu-Juaben district. The results indicated that SVM can improve the classification of oil palm mapping. The estimated area covered by oil palm was 904.95 ha and 993.78 ha for MLC and SVM respectively. SVM and MLC varied in their ability to map and quantify oil palm. SVM is more accurate than MLC. SVM is suitable method for identifying and mapping oil palm.

Ustuner and Barnali, (2015) carried out an application of Support Vector Machines for Landuse Classification Using High-Resolution RapidEye Images: A Sensitivity Analysis. The objective of their study was to investigate the sensitivity of SVM architecture including internal parameters and kernel types on landuse classification accuracy of RapidEye imagery for the study area in Turkey. Four types of kernels (linear, polynomial, radial basis function, and sigmoid) they are used for the SVM classification. A total of 63 different models were developed and implemented for sensitivity analysis of SVM architecture. The traditional Maximum Likelihood Classification (MLC) method was also performed for comparison. The classification accuracies of the best model for each kernel type and MLC are 85.63%, 83.94%, 83.94%, 83.82% and 81.64% for polynomial, linear, radial basis function, sigmoid kernels and MLC, respectively. The results suggest that the choice of model parameters and kernel types play an important role on SVMs classification accuracy. Best model of polynomial kernel outperformed all SVMs models and gave the highest classification accuracy of 85.63% with RapidEye imagery.

Gap: The primary objective of this research was assessing the applicability and sensitivity of SVM in classifying RapidEye images to enhance the agricultural landuse database for the Ministry of Food, Agriculture and Livestock in Turkey. The specific task of SVM sensitivity included analysis of internal parameters and kernel types on landuse classification accuracy. They concluded that Rapid Eye imagery gives satisfactory results for classification on agricultural applications and SVMs can accomplish high classification accuracy with a small size of training datasets. SVM models in all selected cases (best per class for each of their models: linear, polynomial, radial basis function, and sigmoid) outperformed the traditional MLC method.

Melgani, and Bruzzone, (2004) addressed the problem of the classification of hyperspectral remote sensing images by support vector machines (SVMs). First, they propose a theoretical discussion and experimental analysis aimed at understanding and assessing the potentialities of SVM classifiers in hyper-dimensional feature spaces. Then, they assess the effectiveness of SVMs with respect to conventional feature-reduction-based approaches and their performances in hyper-subspaces of various dimensionalities. To sustain such an analysis, the performances of SVMs were compared with those of two other nonparametric classifiers (i.e., radial basis function neural networks and the K-nearest neighbour classifier).

Finally, they studied the potentially critical issue of applying binary SVMs to multiclass problems in hyperspectral data. In particular, their different multiclass strategies are analysed and compared: the one-against-all (OAA), the one-against-one (OAO), and two hierarchical tree-based strategies(BHT-BB). Different performance indicators have been used to support their experimental studies in a detailed and accurate way, i.e., the classification accuracy, the computational time, the stability to parameter setting, and the complexity of the multiclass architecture. The results obtained on a real Airborne Visible/Infrared Imaging Spectro-radiometer hyperspectral dataset shows that any of the multiclass strategy adopted, SVMs are a valid and effective alternative to conventional pattern recognition approaches (feature-reduction procedures combined with a classification method) for the classification of hyperspectral remote sensing data.

The experimental analysis was organized into three main experiments. The first aims at analyzing the effectiveness of SVMs in classifying hyperspectral images directly in the original hyper-dimensional feature space. A comparison with two other nonparametric classifiers is provided as well as an assessment of the stability of these three classification methods versus the setting of their parameters. In the second experiment, SVMs are compared with the classical approach adopted for hyperspectral data classification that is a conventional pattern recognition system made up of a classification method combined with a feature-reduction technique. In these two experiments, they adopted the most popular multiclass strategy used for SVMs that is, the OAA strategy. Finally, the third experiment aims at analyzing and comparing the effectiveness of the different multiclass strategies described in the previous section that is the OAA, OAO, BHT-BB, and BHT-OAA strategies.

Gap: They addressed the problem of the classification of hyperspectral remote sensing data using support vector machines. In order to assess the effectiveness of this promising classification methodology, they considered two main objectives. The first was aimed at assessing the properties of SVMs in hyper-dimensional spaces and hyper-subspaces of various dimensionalities.

In this context, the results obtained on the considered dataset allow to identify the following three properties: (1) SVMs are much more effective than other conventional nonparametric classifiers (i.e., the RBF neural networks and the K-NN classifier) in terms of classification accuracy, computational time, and stability to parameter setting; (2) SVMs seem more effective than the traditional pattern recognition approach, which is based on the combination of a feature extraction/selection procedure and a conventional classifier, as implemented in theirwork; and (3)SVMs exhibit low sensitivity to the Hughes phenomenon, resulting in an excellent approach to avoid the usually time-consuming phase required by any feature-reduction method. Indeed, as shown in the experiments, the improvement in accuracy obtained on the considered dataset by combining SVMs with a feature-reduction technique is definitely insufficient to justify the use of the latter. The second objective of their work concerned the assessment of the effectiveness of strategies based on ensembles of binary SVMs used to solve multiclass problems in hyperspectral data.

Durgesh and Bhambhu (2009) carried out experiment of Data Classification using Support Vector Machine thus showing acomparative results using different kernel functions for all data samples and RSES data sets. In these experiments, they used both methods on different data set. Firstly, they used LIBSVM with different kernel linear, polynomial, sigmoid and RBF kernel was employed. There were two parameters, the RBF kernel parameter ƒÁ and the cost parameter C, to be a set. All the three data sets, diabetes, heart, and satellite, are from the machine learning repository collection. In these experiments, 5-fold cross validation was conducted to determine the best value of different parameter C and ƒÁ. The result shows that the combinations of (C, ƒÁ) is the most appropriate for the given data classification problem with respect to prediction accuracy. Secondly, RSES Tool set was used for data classification with all data set using different classifier technique as Rule Based classifier, Rule Based classifier with Discretization, KNN classifier and LTF (Local Transfer Function) Classifier. The hardware platform used in the experiments was a workstation with Pentium-IV-1GHz CPU, 256MB RAM, and the Windows XP (using MS-DOS Prompt). The Total execution time for all data to predict the accuracy in seconds was also computed.

Gap: They have shown their comparative results using different kernel functions. The experiment results are encouraging. It can be seen that the choice of kernel function and best value of parameters for particular kernel is critical for a given amount of data. However, this experiment was not tested using Medium or high resolution satellite imageries to map urban-rural objects.

Koetza, Morsdorfa, Curtb, Van der Lindenc, Borgnietb, Odermatta, Alleaumeb, Lampinb, Jappiotb and Allgowerd (2006) accessed Fusion of Imaging Spectrometer and Lidar Data Using Support Vector Machines for Land Cover Classification in the Context of Forest Fire Management. The covered site comprised typical Mediterranean vegetation intermixed with urban structures. The research employed remote sensing systems the LiDAR (Optech, ALTM3100) and the imaging spectrometer (AISA-Eagle) were mounted together with a very high-resolution photogrammetric camera (40 cm spatial resolution) on a helicopter and operated by the company HELIOGS. The airborne survey was organized to cover a region of about 13.6 x 3.6 km in a spatial resolution of 1 meter. In the presented study only subset is presented. A combination of the two remote sensing systems, imaging spectrometry (IS) and Light Detection and Ranging (LiDAR), is well suited to map fuel types, especially within the complex wildland urban interface. LiDAR observations sample the spatial information dimension describing geometric surface properties. Imaging spectrometry on the other hand samples the spectral dimension, which is sensitive for discrimination of species and surface types. As a non-parametric classifier Support Vector Machines (SVM) are particularly well adapted to classify data of high dimensionality and from multiple-sources as proposed in this work. The presented approach achieves an improved land cover mapping based on a single SVM classifier combining the spectral and spatial information dimensions provided by imaging spectrometry and LiDAR.

After pre-processing, LiDAR derivatives and spectral bands of the imaging spectrometer were co-registered and as layer stack jointly considered for the classification. Parallel to the airborne survey a comprehensive field campaign was conducted for the validation of the fuel types derived from the observations of the two remote sensing systems. The field measurements collected describe the relevant fuel types, including a specific characterization of the WUI interface and species composition. Fuel properties such as biomass and fuel moisture content were also sampled. The land cover classification was performed by the non- parametric Support Vector Machines (SVM). The SVM classification was trained and applied to three different data sets representing different input sources and information dimensionality.

Gap: The method presented in this study is capable of a joint one-step SVM classification for the fusion of multiple-source remote sensing data provided by an imaging spectrometer and a LiDAR. The SVM classifier was able to efficiently exploit the significantly increased information content in the (hyper) spectral and the three-dimensional dimension. Specifically, the SVM generalized well, even when only small training sets were available for the classification of the high dimensional data provided the multiple data sources. The three- dimensional information of the LiDAR data complemented well the spectral information leading to a significant increase in the overall land cover classification accuracy relative to the pure spectral information input. Important features of fuel types as the vertical structure vegetation and houses could be assessed with higher accuracy and reliability. This enhanced mapping of the wild land urban interface can be a significant input to forest fire behaviour models leading to improved risk assessment and mitigation of forest fires.

Tarun (2014) carried out a Supervised Classification of Remote Sensed data using Support Vector Machine. Two different formats of remote sensed data were considered for the same. The first format is a comma separated value (CSV) format wherein a classification model is developed to predict whether a specific bird species belongs to Darjeeling area or any other region. The second format used is raster format which contains image of Andhra Pradesh state in India.

The data under consideration were first pre-processed by Jeyanthi (2007). In the case of CSV datasets comprising of information of birds of North-east India the attributes considered are id, family, genus, specific_ epithet, latitude, longitude, verbatim _scientific_ name, verbatim_ family, verbatim_ genus, verbatim_ specific_ epithet and locality. A variable called churn acts as a class label which would categorize the data into two categories viz one having data sets of birds from Darjeeling area and the other having data sets of birds belonging to other north eastern parts in India. Before applying the classification, the data sets are cleaned to remove any missing values. In the case of raster data set, a TIFF image was used. The image comprises of a map of Andhra Pradesh, a state in India. Initially a region of interest (ROI) wascaptured and later supervised SVM classification methodology was applied.

Support vector machine classification method is used herein to classify the raster image into categories. One category represents land and the other water wherein green colour is used to represent land and light blue colour is used to represent water. Later the classifier is evaluated using kappa statistics and accuracy parameters.

Gap: the SVM classification method was used to build a classification model for two datasets. The first data set is of CSV format and the second one is a raster TIFF image. Later the classification model is validated against a test data set which is a subset of the input dataset. The performance of SVM is calculated using kappa statistics and accuracy parameters and it is established that for the given data sets SVM classifies the raster image dataset with better accuracy than the CSV dataset.

Pontil and Verri, (1998) used linear SVMs for 3D object recognition. They illustrate the potential of SVMs on a database of 7,200 images of 100 different objects. The proposed system does not require feature extraction and performs recognition on images regarded as points of a space of high dimension without estimating pose. The excellent recognition rates achieved in all the performed experiments indicate that SVMs are well-suited for aspect-based recognition.In order to verify the effectiveness and robustness of the proposed recognition system, they performed experiments on the COIL images under increasingly difficult conditions.

They first considered the images exactly as extracted from the COIL database, then added pixel-wise random noise, bias in the registration of the test images, and a combination of the two. Finally, they studied the sensitivity of the system to moderate amount of occlusions and compared the obtained results against a simple perceptron.

Gap: They have assessed the potential of linear SVMs in the problem of recognizing 3D objects from a single view. As shown by the comparison with other techniques, it appears that SVMs can be effectively trained even if the number of examples is much lower than the dimensionality of the object space. This agrees with the theoretical expectation that can be derived by means of VC-dimension considerations (Vapnik, 1995). The remarkably good results which they have reported indicate that SVMs are likely to be very useful for direct 3D object recognition.

Xie, (2006) developed aSupport Vector Machines method for land use change modelling with the capacity to effectively address land use change data, which might be a mixture of continuous and categorical variables that are not normally distributed. A SVMs land use change modelling framework was developed using C++ programming language. The modelling framework was integrated into ArcMap as an extension so as to make use of ArcMap's powerful spatial data processing and visualization capacity. Support vector machines were applied to model land use change in Calgary, Canada from 1985 to 2001. Calgary is located in southern Alberta on the eastern edge of the Rocky Mountain Foothills at the merging of the Bow and Elbow rivers. Calgary covers an area of about 720 square kilometers and thus a 25 km x 35 km rectangle should be sufficient to include it. Some special issues regarding the implementation of SVMs (e.g. regularization parameter selection, kernel selection, vector normalization) are carefully studied. Land use change data was generally unbalanced, which leads to low performance of classifying the minority, namely, the changed land parcels. Improvement to standard SVMs was introduced to achieve uniform performance for both the changed and unchanged data. A secondary improvement was also implemented which increases robustness and allows SVMs to cope with outliers in the land use change dataset thus providing reliable performance when there is a certain level of disturbance on land use change caused by other unobserved factors.

Gap: To investigate the performance of SVMs on land use change modeling, a SVMs land use change modeling framework was implemented and applied to a case study of modeling land use change in Calgary from 1985 to 2001. The performance of SVMs was compared with that of a well-studied land use modeling approach, namely, spatial logistic regression(SLR), the comparison showed that SVMs were superior to SLR.However, the method of error matrix and kappa testing was not mentioned.

Two improvements of standard SVMs were developed to tailor SVMs to better fit the characteristics and requirements of land use change modeling. The first improvement aimed to improve the accuracy of classifying smaller class when the training set was unbalanced. By changing the objective function and giving different weights for positive and negative data, the improvement proved to be effective in providing uniform high performance for unbalanced land use change data. The other improvement aimed to improve the robustness of SVMs especially in the case of unbalanced data. A robust SVMs algorithm that detected outliers and removed them from the support vectors was introduced and tested. The result showed that the robust SVMs could efficiently improve robustness.

Shi and Yang, (2012) examined the utilities of support vector machines (SVM) as a promising pattern recognition technique for landscape mapping particular for heterogeneous areas. It was organized into two major parts, beginning with a brief introduction of some basic knowledge on SVM and a review on the research status and possible challenges of using SVM for landscape mapping. The review focuses on some comparative studies that demonstrated the effectiveness of SVM over other conventional classifiers. Based on the review, they further discuss several areas that need additional research in order to improve SVM classification accuracies and reduce computational burdens, which are mostly related to appropriate treatments of some parametric and nonparametric factors. The second part of the paper discusses their implementation of SVM to map various land cover types from a remote sensor image covering an urban area, demonstrating the robustness of this type of pattern recognition technique for mapping heterogeneous landscapes.

The study site covers the entire Gwinnett County, a suburban county located at north eastern Atlanta metropolitan area, Georgia, USA. Its landscape is characterized by a mosaic of complex land use and land cover types, and therefore Gwinnett is an ideal site to examine the effectiveness of SVM for heterogeneous landscape mapping.A cloud-free Landsat-5 Thematic Mapper (TM) image dated on 19 May 2007 was acquired from USGS EROS Data Centre, and a subset of this scene covering the entire Gwinnett County was actually used in their study. The image has been geometrically corrected at the EROS data centre, and no further pre-processing was conducted. The spatial resolution of this image is 30 m for all six non-thermal infrared bands, and 120 m for the thermal band. It was projected into the Universal Transverse Mercator Zone 16N with NAD 83 as the horizontal datum.

They designed a land use/cover classification scheme based on the Anderson scheme (Anderson et al., 1976) and their field surveys across the Atlanta metropolitan area. The study area covers a mosaic of different land use cover types, and their classification system includes ten major categories: high-density urban, low-density urban, barren or fallow land, pasture and cropland, grassland, shrub and scrub, evergreen forest, deciduous forest, mixed forest, and water.After the classification scheme was adopted, they carefully selected training samples for each of the ten major categories by using several reference sources such as the high-resolution images from Google Earth and the 2006 National Land Cover Data (NLCD). For the information classes with multiple spectral classes, they collected at least one training set with 25-35 pixels for each spectral class.MLC also misclassified some evergreen forest patches into water, barren land patches into high density urban, and grassland patches into low density urban and cropland. Contrastingly, SVM seemed to have done a better job in mapping spatially scattered patches. And SVM had correctly classified the residential patches on all the three sites and the pasture patches on Site

Gap: They have reviewed the research status of using support vector machines (SVM) for landscape mapping with special attention on heterogeneous landscape types. Then, they have implemented this technique to map various land cover types in an urban area from a remote sensor image. Their studies further confirm that SVM can significantly outperform the maximum likelihood classifier (MLC), the most widely used pattern recognition method in the remote sensing community. They found that SVM can significantly improve mapping accuracy, particularly for spectrally and spatially complex landscape categories. However, the study fails to indicate how the software or the classifier was developed or complied.

Bekkari, (2012) proposed a methodology in SVM classification of high resolution urban satellites images using Haralick features. The SVM classification was conducted using a combination of multi-spectral features and Haralick texture features as data source. They have used homogeneity, contrast, correlation, entropy and local homogeneity, which were the best texture features to improve the classification algorithm. The result was compared with both a standard SVM classifier and a SVM classifier with a Graph Cuts approach that introduces spatial domain information applied as a post-classification. The proposed approach was tested on common scenes of urban imagery. Results showed that SVMs, especially with the use of Haralick texture features, outperform the SVM classifier with post-processing in term of the global accuracy. The experimental results indicate a mean accuracy value of 94.045 % which is very promising.

Gap: They have presented a method which makes it possible to combine spatial and spectral information by the use of Haralick texture features to refine the classification of multispectral satellite images. The experimental results are promising and comparisons with Graph-Cuts approach applied to the spectral classification have shown that the proposed method is able to find better classes; however, it remains to improve even more of these results. The workflow of this study can be used in other remote sensing application, especially, in rural areas for thematic land cover. However, they did not study the different kernel choice in order to determine the appropriate one, for this type of image classification.

Zhang, Lin, and Ning (2013) noted that the Object-based point cloud analysis (OBPA) is useful for information extraction from airborne LiDAR point clouds while carrying their research on “SVM-Based Classification of Segmented Airborne LiDAR Point Clouds in Urban Areas” hence, proposed an object-based classification method for classifying the airborne LiDAR point clouds in urban areas. In the process of classification, the surface growing algorithm was employed to make clustering of the point clouds without outliers, thirteen features of the geometry, radiometry, topology and echo characteristics are calculated, a support vector machine (SVM) was utilized to classify the segments, and connected component analysis for 3D point clouds was proposed to optimize the original classification results. Three datasets with different point densities and complexities were employed to test their method. Experiments suggest that the proposed method is capable of making a classification of the urban point clouds with the overall classification accuracy larger than 92.34% and the Kappa coefficient larger than 0.8638, and the classification accuracy was promoted with the increasing of the point density, which is meaningful for various types of applications.

Gap: Airborne LiDAR point cloud classification is meaningful for various applications, but it is a challenging task due to its complexity. In this research, classification was based upon object-based point cloud analysis, which defines the general two steps: clustering and classification. Particularly, surface growing algorithm is employed to make a clustering, and SVM was used to classify the points in a segment-wise style. There are two main contributions of the work: (1) the use of SVM to classify surface growing segmented objects (OBPA approach); (2) the proposed connected-component analysis for 3D point clouds. The experimental results suggest that their method is capable of making a good coarse urban scene classification.

Bellman and Shortis, (2014)used a set of classification test data, in the form of square image patches, was extracted from three frames of colour aerial photography. The photographs were originally acquired for a project over the city of Ballarat in Victoria. The photographs were recorded at a scale of 1:4000 and had been scanned on a photogrammetric scanner at a resolution of 15 microns. Each image patch was 256 by 256 pixels and contained either a single building or a non-building area of the image. Although some care was taken to centre the building within the image patch, the exact location of the building in the image patch varied. The orientation of the building within the image patches also varied. This leads to a broader representation of the building class than if the buildings were carefully aligned in each image patch but created a more difficult classification problem.

The classification test was based on a balanced test set of 100 building images and 100 non-building images. Image coefficients were extracted using the wavelet process described earlier. A public domain Support Vector Machine (Joachims1998) was used to classify the image patches into building or non-building categories. The image patches used to train the SVM classifier using several different kernels including polynomial and sigmoidal kernels.

However, the best results were obtained with a simple linear kernel with no bias. Of the two hundred image patches, only one patch (of a large swimming pool) was classified incorrectly. Although this result appeared to be very good, the confidence measures produced by the SVM training suggested a reliability of only 55%. This could be due to over fitting of the decision surface to the data. However, the reliability measures produced by the SVM are also known to be pessimistic (Joachims,1998), due to the unbounded nature of the problem. To investigate this further, an extensive leave-one-out test was undertaken. This produced a revised reliability measure of 73%. Although the reliability estimate improved, this indicates there may still be some over fitting of the data. As a further test, 20 building image patches were withheld from the training data and the SVM was re-trained. The withheld patches were then classified by the new SVM. Of the 20 building patches, only 8 were classified as buildings. This result is similar to the original reliability estimate. In this case, the decision surface of the SVM is unlikely to generalize well to a broader set of data. This could be due to the small size of the training set.

Other resolutions were tried but at higher resolutions, the number of coefficients expands rapidly and the lower resolutions did retain sufficient information about the image to enable classification. One limitation of this implementation is the size of the coefficient data set produced from the wavelet transform. As the wavelet transform is over-sampled, each image patch generates 960 coefficients. The current algorithm does not attempt to optimize the storage of these coefficients.

Gap: Most computer vision problems require salient features to be detected in the image as part of the processing strategy. In this application, a simple form of wavelet analysis was used. The images are characterized using the wavelet analysis to provide multi- resolution data for the machine learning phase. The learning in this situation was complicated by the high dimensionality of the image data and its wavelet derivatives. One limitation of this implementation is the size of the coefficient data set produced from the wavelet transform. As the wavelet transform is over-sampled, each image patch generates 960 coefficients. The current algorithm does not attempt to optimize the storage of these coefficients.The small size of the training set should be increased in order to boost the generalization of broader dataset.

2.2.4 Summary of gaps identified in the literature

From the related literatures reviewed, the following findings were observed. Many of the previous researches used the traditional statistical methods, e.g. Markov chain analysis, Maximum Likelihood classifiers, multiple regression analysis, principal component analysis, and logistic regression. However, their drawbacks limit their efficiency in land use modeling, (Xie, 2006). Some studies were done in area of satellite image classification using different types of self-developed data mining algorithms and machine learning approach except SVM classifier, while using different kinds of datasets. Some researchers applied the use of mostly integrated SVM with modified SVM in their experiments to assess the development in other field outside remote sensing and satellite surveying, for example, Liu, Chang and Ma (2014) carried out the assessment of groundwater quality based on Support Vector Machine. Studies by some researchers were based on self-developed SVM e.g Clustering-Based SVM (CB-SVM) and the results tested with other machine learning algorithms for various object classification., while the remaining researchers used only MatLab complied SVM algorithm for object classification and few integrated SVM with other algorithm like Maximum likelihood classifier in purely hyper-spectral and high resolution satellite image classification with various methodology for example, supervised or unsupervised, object based or pixel based, among others.

The application of the statistical learning theory of SVM for mapping and surveying were at a developmental and testing stage. Nevertheless, SVM is now gradually leaving the validation stage to confirmation stage for global acceptability and thus, very fewresearchers used commercial packaged SVM to carry out their analysis.However, their major drawbacks limit the use of the SVM for mapping and solving the real-world problems using data acquired from earth observation sensors and other remote sensing techniques.Some researchers validated the results of their image classification using Orthophoto-derived vector map of their study area using software-based SVM, but however fail to test the accuracy of classification using the same software-based statistical tools like Confusion matrix and kappa’s index.

With these reviews, it is still obvious that few researches have been done in classification of surface features in high resolution satellite imageries, using SVMespecially in Africa and Nigeria in particular. Hitherto, there is a limited prior research and inconsistency of research results in this study area, especially in the LU/LC map produced, this is because, the few image classifications that were done, were classified without considering the whole extent of Nnewi-north local government area, thus creating a gap.

Table: 2.2. Summary of Literature Reviewed and Gaps Identified					
S/N	**AUTHOR (S)**	**RESEARCH TOPIC**	**METHODOLOLGY AND RESULTS**	**ALGORITHM (S) USED**	**GAPS**
SECTION I Data Mining and Machine Learning Algorithm for Various Classifications					
1	Moussa and El-Sheimy (2002)	Man-Made Objects Classification from Satellite/Aerial Imagery Using Neural Networks	Their work uses a supervised neural network in which a set of inputs and associated outputs were provided to train the network at the initial stage. The training of the neural network uses the calculated descriptors for man-made objects from previously classified images. After the training phase, the descriptors of the test images objects are fed into the neural network to perform the classification. The results of the proposed approach were presented based on a reference solution for evaluation purposes.	Neural Networks	The separabilties of those regions of interest and classified results accuracy were not validated because the method used was not mentioned.
2	Sitaram and Manjunath(2006)	Modelling and Detection of Geospatial Objects Using Texture Motifs	They demonstrated the model training and selection methodology for different objects given a limited dataset of each. Then, emphasize the utility of such models for detecting the presence or absence of geospatial objects in large aerial image datasets comprising tens of thousands of image tiles.	Gabor filter and Gaussian mixture model (GMM)	They demonstrated the use of texture Motifs for modelling geospatial objects using Gabor filter and Gaussian mixture model (GMM).
3	Khoshelham, et al (2010)	Performance evaluation of automated approaches to building detection in multi-stheirce aerial data, Ancona, Italy.	They undertake a comparative evaluation of three common data fusion and classification methods, namely Bayesian, Dempster–Shafer and AdaBoost, as applied to the detection of buildings in multi-source aerial data. They compare the performance of the fusion methods with the less elaborate method of thresholding, the normalized DSM. Table 2.2. Continued…	Bayesian, Dempster–Shafer and AdaBoost	Errors were found at building boundaries and in areas where dense trees were present. There is need to use satellite remote sensing data of higher resolution to crosscheck this finding.
4	Bruzzone and Carlin	A Multilevel Context-Based	They carried out a multilevel context-based system for classification of very high spatial resolution images (panchromatic and	SVMs module, Gaussian pyramidand	The SVMs classifier used are authors developed

	(2006)	System for Classification of Very High Spatial Resolution Images	pan-sharpened). They come up with a novel pixel-based system for the supervised classification of very high geometrical resolution images. This system is aimed at obtaining accurate and reliable maps both by preserving the geometrical details in the images and by properly considering the spatial context information. Based on hierarchical segmentation with constraints.	Gaussian decomposit or	algorithms which required further validation by other remote sensing scientist. Their result was considered as a pre-processing stage aimed at driving the feature-extraction phase. In this case, the use of the geometrical and relational information does not increase the separability.
5	Bayoudh, et al (2012)	Automatic Learning of Structural Knowledge from Geographic Information for updating Land Cover Maps	Their work consists in implementing machine learning methods for structural and symbolic knowledge extraction from land use/cover maps and from various complementary geographic information layers. Table 2.2. Continued…	Aleph, an Inductive Logic Programming (ILP) system.	Aleph, an Inductive Logic Programming (ILP) system developed required further validation by other remote sensing scientist.
6	Singh, et al (2015)	Land Information Extraction with Boundary Preservation for High Resolution Satellite Image	They developed an efficient technique for edge detection to define land boundaries and feature selection technique for land information extraction. So, they succeeded in using an edge detection technique and object based classification to extract the land information automatically and then associate the area detail with each land object. In this method they modify the Sobel operator to detect gradient at horizontal and vertical level.	A modify Sobel operator using the segmentation technique	A modify Sobel was operated using the segmentation technique and thus required a comparative advantage analysis.

Table 2.2. Continued…

7	Arun and Katiyar (2013)	An Intelligent Approach towards Automatic Shape Modelling and Object Extraction from Satellite Images using Cellular Automata Based Algorithm	They proposed a frame work for accurate detection of features along with automatic interpolation, and interpretation by modelling feature shape as well as contextual knowledge using advanced techniques such as SVRF, Cellular Neural Network, Corset, and MACA. Developed methodology has been compared with contemporary methods using different statistical measures. Investigations over various satellite images revealed that considerable success was achieved with the CNN approach. CNN has been effective in modelling different complex features effectively and complexity of the approach has been considerably reduced using corset optimization. The system has dynamically used spectral and spatial information from the satellite image of Bhopal and Chandrapur cities of India.	SVRF, Cellular Neural Network, Corset Optimization, and MACA in ERDAS	The dataset were from various sources which include Google Earth image, Differential Global Positioning System (DGPS), PanSensor image and LANDSAT image required resampling which was not stated. And it is important to compare the results using Error Matrix and Kappa index other than the statistical method mentioned.
8	Ohlhof, et al (2015)	Semi-Automatic Extraction of Line and Area Features from Aerial and Satellite Images	Their project was carried out by two software companies ESG and Inpho, were they developed an operational system for the semi-automatic extraction of line and area features in 2D and 3D based on an existing software platform. The complete system was delivered to the Geo-Information Office (AGeoBw) of the German Federal Armed Forces and has [illegible] May 200[illegible] [illegible]el 1 data and the generation of the military basic vector database.	inJECT software extended with new automation modules and a strong GIS interface, Dynamo and Geo-media.	inJECT software requires comparative analysis with the recently developed software to verify this claims.
9	Iovan, et al (2008)	Detection, Characterization and Modelling Vegetation in Urban Areas from High	They presented a complete image analysis system which, from high-resolution colour infrared (CIR) digital images, and a Digital Surface Model (DSM), extracts. Segments and classifies vegetation in high density urban areas, with very high reliability.	A supervised classification system based on a Support Vector Machines	The research should be tested using a large-scale dataset, or using supplementary information from

		Resolution Aerial Imagery	The process starts with the extraction of all vegetation areas using a supervised classification system based on a Support Vector Machines (SVM) classifier. The result of this first step is further on used to separate trees from lawns using texture criteria computed on the DSM. Tree crown borders are identified through a robust region growing algorithm based on tree - shape criteria.	(SVM) classifier	hyperspectral or LiDAR.
10	Do, et al (2005)	A Simple, Fast Support Vector Machine Algorithm for Data Mining	They created a new algorithm that is very fast for building incremental, parallel and distributed SVM classifiers. SVM and kernel related methods have shown to build accurate models but the learning task usually needs a quadratic programming, so they observed the learning task for large datasets requires big memory capacity and a long time. Their result presented a new SVM algorithm being able to deal with very large datasets in linear and nonlinear classification tasks on PCs. They have extended the recent finite Stepless Newton SVM algorithm to build incremental, parallel and distributed SVM.	A new SVM algorithm , Newton SVM algorithm	The algorithm developed need to be tested using real-world satellite image dataset.
11	Liu, et al (2009)	Groundwater qualityassessment based onSVM	They researchers used SVM method to assess karst groundwater quality at the Niangziguan fountain region of Haihe River basin. Eight assessment factors were selected to randomly generate sample set. All test samples were classified correctly after training the model. The result shows that such a method solves the complex nonlinear relationship between assessment factor and water quality grade. It offers high pre[illegible] is a rea[illegible]sment method.	SVM	These methods do not solve the complex nonlinear relationship between assessment factor and quality grade, and the assessment result can be largely affected by the subjective factors of the assessing person.
SECTION II					

Table 2.2. Continued…

		SVM Application in other Field			
12	Decherchi, et al (2009)	Using Unsupervised Analysis to Constrain Generalization Bounds for Support Vector Classifiers	They developed a general method to compute the generalization bounds for SVMs, which is based on referring the SVM parameters to an unsupervised solution, and shows that such an approach yields tight bounds and attains effective model selection. When one estimates the generalization error, one uses an unsupervised reference to constrain the complexity of the learning machine, thereby possibly decreasing sharply the number of admissible hypothesis. They adopted Vector Quantization (VQ) as a representation paradigm, and introduce a biased regularization approach in bound computation and learning. Experimental results validate the proposed method on complex real-world data sets.	Vector Quantization (VQ) in SVM	The generalization bounds parameters should be also tested using supervised solution and the approach can be extended to other classifier models.
13	Huysmans, et al (2006)	Country Corruption Analysis with Self Organizing Maps and Support Vector Machines	They integrated SOM and SVM and observed that the LS-SVM model provides a better forecasting accuracy than the corresponding linear model. However, they noted that the LS-SVM has a serious disadvantage: it is very difficult to understand the motivation behind this model's decisions due to its complexity. To relieve this opacity restriction and to gain more insight in Table 2.2. Continued… n d g [illegible] errors onto a self-organizing map	They integrated SOM, SVM and LS-SVM model	This ideology should be evaluated using Remote sensing data and applied in various remote sensing applications.
14	L'ubor-Ladický, et al (2012)	Latent SVMs for Human Detection with a Locally Affine Deformation Field	They proposed a latent deformable template model with a locally affine deformation field, which allows for more general and more natural deformations of the template while not over-fitting the data; and also provide a novel inference method for this kind of problem. The deformation model was to measure the distances between training samples, and their result show how this can be used to cluster the problem into several modes,	Latent SVMs	The random field formulated with the locally affine constraints could be used for other computer vision tasks and in infrastructure deformation mapping and modelling.

			corresponding to different types of objects, viewpoints or poses. Their method leads to a significant improvement over the state-of-the-art with small computational overhead. The training samples are clustered into 10 models.		
15	Aymerich, et al (2010)	Potential Support Vector Machines for phytoplankton fluorescence spectra classification: Comparison with Self - Organizing Maps	These researchers presented a comparison between two different techniques for fast phytoplankton discrimination: Self-Organizing Maps (SOM) and Potential Support Vector Machines (P-SVM), evaluating its capability to achieve phytoplankton classification from its fluorescence spectra. Five different cultures representing the major algae divisions were selected and grown under the same conditions. Excitation-Emission Matrices were acquired every day. SOM method was used as a first approach to phytoplankton discrimination from excitation and emission fluorescence spectra. Herein they present the results using P-SVM, and they compare the results using both techniques.	Self-Organizing Maps (SOM) and Potential Support Vector Machines (P-SVM),	Self-Organizing Maps (SOM) and Potential Support Vector Machines (P-SVM) should be evaluated using Remote sensing data.
16	Song, et al (2002)	Robust Support Vector Machine with Bullet hole Image Classification	They proposed a general robust SVM algorithm against outliers. By adding the distance between each data point and the centre of classes to form the margin of separating hyperplane, good robust performance is achieved. The simulation of robust SVM with different kernel functions and regularization parameters has been presented to show that the robust algorithm can be used for pattern classification problems with different difficulty levels. The experiment results show that the decision boundary becomes less dithered and the number of support vectors of the ro Table 2.2. Continued… icantly cc andard SVM.	robust SVM with different kernel functions	Potential of robust SVM with different kernel functions should be evaluated using Remote sensing data in solving real-world problem.
17	Huang, (2011)	Using Genetic Algorithm	He adopts genetic algorithm for the search of the best SVM parameter and the selection of the best SVM	SVM model, Genetic	Appropriateness of the parameters and

		Optimization SVM to Construction of Investment Model	prediction variable, and was compared with Logistic Regression for the classification prediction capability of stock investment. From the classification prediction result and the result of AUC of the models presented, it can be seen that the SVM after adjustment of input variables and parameters have classification prediction capability relatively superior to that of the other three models.	algorithm and Logistic Regression model.	prediction variable can affect prediction capability of the entire model. This approachshould be evaluated using Remote sensing data in prediction of urban landuse and landcover variation.
18	Anuja-Kumari and Chitra (2013)	Classification of Diabetes Disease Using Support Vector Machine	They have used datasets for diabetes disease from the machine learning laboratory at University of California, Irvine. All the patients" data are trained by using SVM. The choice of best value of parameters for particular kernel is critical for a given amount of data SVM approach can be successfully used to detect a common disease with simple clinical measurements, without laboratory tests. In the proposed work, SVM with Radial basis function kernel is used for classification. The performance parameters such as the classification accuracy, sensitivity, a [illegible] h [illegible] Table 2.2. Continued… a good option for the classification process	Matlab R2010a, Receiver Operating Characteristic (ROC) curve, SVM with Radial basis function kernel was used for classification	The result of the SVM classifier can be improved by using feature subset selection process.This approach should be evaluated using Remote sensing data in prediction of urban landuse and landcover variation.
19	Anis and Abdelaziz, (2012)	Complex Support Vector Machine Regression for Robust Channel Estimation in LTE DOWNLINK System''	They describe a nonlinear complex SVR based channel estimation technique for a downlink LTE system in the presence of Gaussian noise and impulse noise interfering with OFDM pilot symbols in high mobility environment. The proposed method was based on learning process that uses training sequence to estimate the channel variations. Their formulation was based on nonlinear complex SVR specifically developed for reference-based OFDMA systems. Simulations have confirmed the capabilities of the proposed	Nonlinear complex Support Vector Machine Regression (SVR) and Long Term Evolution (LTE) downlink.	This approach should be verified using Remote sensing data in prediction of urban landuse and landcover variation or in testing and enhancing the global positioning system (GPS) signal.

			nonlinear complex OFDM-SVR estimator in the presence of Gaussian and non-Gaussian impulse noise interfering with the reference symbols for a high mobile speed when compared to conventional LS and Decision feedback methods for a highly selective time varying multipath fading channel.		
			SECTION III *SVMs and other Algorithms for ObjectsClassification*		
20	Yu, et al (2003)	Classifying Large Data Sets Using SVMs with Hierarchical Cluster	They proposed a new method called CB-SVM (Clustering-Based SVM) that integrates a scalable clustering method with an SVM method and effectively runs SVMs for very large data sets. CB-SVM applies a hierarchical micro-clustering algorithm that scans the entire data set only once to provide an SVM with high quality samples that carry the statistical summaries of the data such that the summaries maximize the benefit of learning the SVM. CB-SVM tries to generate the best SVM boundary for very large data sets given limited amount of resources. Their experiments on synthetic and [illegible] SVM is [illegible] Table 2.2. Continued… [illegible]ge data sets while also generating high classification accuracy.	a new method called CB-SVM (Clustering-Based SVM) and a SVM method	The separabilties of those data sets and classified results accuracy were not validated using error matrix or Kappa because the method was not mentioned.
21	Jin, et al (2011)	Accurate urban road model reconstruction from high resolution remotely sensed imagery based on Support Vector Machine and Gabor filters.	Carried an automatic road feature extraction approach and applied it to a number of aerial images in urban areas. Supervised SVM image classification technique was employed to segment the road surface from other ground details, and the road pavement markings are detected on the generated road surface with Gabor filters. Their experiments using several pan-sharpened aerial images of Brisbane, Queensland have validated the proposed method were the objective of the experiment was to determine the performance of the proposed road feature extraction approach	Support Vector Machine and Gabor filter	Supervised SVM classification will be validated using error matrix or Kappa because the statistical testing was not mentioned.

			quantitatively over the study area. A dataset of aerial images located in South Brisbane, Queensland were selected as the study areas.		
22	Yao, et al (2013)	K-SVM: An Effective SVM Algorithm Based on K-means Clustering	Their experiment used a novel approach based on clustering algorithm, in which only a small subset was selected from the original training set to act as their final training set. Their algorithm works to select the most informative samples using K-means clustering algorithm, and the SVM classifier is built through training on those selected samples. Experiments show that their approach greatly reduces the scale of training set, thus effectively saves the training and predicting time of SVM, and at the same time guarantees the generalization performance. Their developed algorithm and LIBSVM are implemented on three datasets, respectively.	SVM classifier and LIBSVM	The algorithm was implemented with Matlab and libsvm toolbox. The kernel function they used here is the Gaussian kernel thus the requires further studies in comparing other related algorithm.
23	Radha and Rehna (2014)	Retrieval of Geographic Images Using Multi Support Vector Machine and Local Invariant Features	They presented an investigation into local invariant features for overhead image retrieval, the first such study of its kind. They demonstrated that local invariant features are more effective than standard features such as colour and texture for image retrieval of LULC classes in high-resolution aerial imagery. They performed image retrieval using Multi Support Vector Machine and it was seen that the false image retrieval has been considerably reduced.	Multi Support Vector Machine	How these results were validated was not stated and hence require such verification.
24	Platt, (1998)	Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines	Proposes a new algorithm for training SVM: Optimization SMO. Training a SVM requires the solution of a very large quadratic programming (QP) optimization problem. SMO breaks this large QP problem into a series of smallest possible QP problems. These small QP problems are solved analytically, which avoids using a time-consuming numerical QP optimization as an inner loop. The	Support vector machines and Sequential Minimal Optimization (SMO).	Sequential Minimal Optimization (SMO) should be tested using in applied Remote sensing and on non-linear dataset.

			amount of memory required for SMO is linear in the training set size, which allows SMO to handle very large training sets. Because matrix computation is avoided, SMO scales somewhere between linear and quadratic in the training set size for various test problems, while the standard chunking SVM algorithm scales somewhere between linear and cubic in the training set size. SMO's computation time is dominated by SVM evaluation; hence SMO is fastest for linear SVMs and sparse data sets. On real-world sparse data sets, SMO can be more than 1000 times faster than the chunking algorithm.		
25	Psaltis and Ioannidis (2010)	Supervised Change Detection On Simulated Data Employing Support Vector Machines	They illustrated a methodology to accomplish supervised change detection on simulated data employing support vector machines The main idea of the proposed procedure follows the supervised classification paradigm. The first step is to layer the available data for the same region in different time periods. Then evaluate a number of predefined cues for the whole region and use some manually collected positive and negative samples to train a classifier. Finally, this classifier can be used to assert change in the remaining data. The base data chosen are very high resolution orthoimages and digital surface models (DSMs) because they offer both the radiometric and geometric information needed for robust change detection. The classifier selected is the support vector machines (SVM) algo Table 2.2. Continued… me sign ver alternative methods.	support vector machine (SVM) and developed algorithm	A high quality DSM can greatly enhance the overall performance of the algorithm developed. Support vector machine used can be commercial package type or authors developed SVM algorithm but this was not stated.
26	Boley and Caoy, (2004)	Training Support Vector Machine using	Noted that the training of support vector machines involves a huge optimization problem and many specially designed algorithms have been proposed by various authors.	ClusterSVM, tested against the popular training	ClusterSVM accelerates the training process by exploiting the

		Adaptive Clustering	And thus, proposed another algorithm called ClusterSVM that accelerates the training process by exploiting the distributional properties of the training data, that is, the natural clustering of the training data and the overall layout of these clusters relative to the decision boundary of support vector machines. The proposed algorithm first partitions the training data into several pair-wise disjoint clusters. Then, the representatives of these clusters are used to train an initial support vector machine, based on which they can approximately identify the support vectors and non-support vectors.	algorithm SMO	distributional properties of the training data but required to evaluated using HRSI covering a large area
27	Suykens , et al (2001)	Weighted Least Squares Support Vector Machines: Robustness And Sparse Approximation	They have shown how to obtain robust estimates within the LS-SVM framework in the case of outliers and heavy tailed non-Gaussian error distributions. This was done by applying a weighted LS-SVM version. While least squares in standard parametric linear regression has a low breakdown point, LS-SVM's with RBF kernel have much better properties. Nevertheless, the robustness can be further enhanced by applying an additional weighted LS-SVM step. While standard SVM's possess a sparseness property for the solution vector, this feature is lost when considering LS-SVM's. However, they have shown how sparseness can be obtained by Table 2.2. Continued… ıe	Least Squares Support Vector Machines (LS-SVM)	A fundamental research is required in this procedure of un-weighted LS-SVM.
28	Mahi, et al (2014)	Zernike Moments and SVM for Shape Classification in Very High Resolution Satellite Images	Their proposed approach comprises three steps. First, the image was segmented into homogeneous objects based on the spectral and spatial information. MeanShift segmentation method was used for this end. Second, a Zernike feature vector was computed for each segment. Finally, a SVM-based classification using the feature vectors as inputs was	Support Vector Machines (SVM) and comparison with ENVI	However, the analysis of the results shows errors lies with confusion between buildings and grass areas. These confusions are

			performed. Experimental results and comparison with ENVI commercial package confirm the effectiveness of the proposed approach.		mainly due to the fact that grass areas and buildings have similar shapes. This problem can be avoided by computing the NDVI.
29	Hsu and Lin (2002)	A Comparison of Methods for Multiclass Support Vector Machines	Studied the Comparison of Methods for Multiclass Support Vector Machines. The SVMs were originally designed for binary classification. How to effectively extend it for multiclass classification is still an on-going research issue. Several methods have been proposed where typically they construct a multiclass classifier by combining several binary classifiers. Some authors also proposed methods that consider all classes at once. Especially for methods solving multiclass SVM in one step, a much larger optimization problem was required so up to now experiments are limited to small data sets. Their experiments indicated that the “one-against-one” and DAG methods were more suitable for practical use than the other methods.	Multiclass Support Vector Machines.	Result shows that for large problems, methods of considering all data at once in general need a further support vector, which meansa much larger optimization problem is required.
SECTION IV Table 2.2. Continued… SVM tion					
30	Ustuner, et al (2015)	Application of Support Vector Machines for Landuse Classification Using High-Resolution RapidEye Images: A Sensitivity Analysis	The primary objective of this research was assessing the applicability and sensitivity of SVM in classifying RapidEye images to enhance the agricultural landuse database for the Ministry of Food, Agriculture and Livestock in Turkey. The specific task of SVM sensitivity included analysis of internal parameters and kernel types on landuse classification accuracy. They concluded that RapidEye imagery gives satisfactory results for classification on agricultural applications and SVMs can accomplish high classification accuracy with a small size of training	Support Vector Machines	the applicability and sensitivity of SVM in classifying RapidEye images to enhance the agricultural landuse can be test on urban landscape to see if it can accomplish high classification accuracy with a large size of

			datasets. SVM models in all selected cases (best per class for each of the models: linear, polynomial, radial basis function, and sigmoid) outperformed the traditional ML method.		training datasets
31	Melgani and Bruzzone (2004)	classification of hyperspectral remote sensing images by support vector machines (SVMs)	The experimental analysis was organized into three main experiments. The first aims at analyzing the effectiveness of SVMs in classifying hyperspectral images directly in the original hyperdimensional feature space. A comparison with two other nonparametric classifiers is provided as well as an assessment of the stability of these three classification methods versus the setting of their parameters. In the second experiment, SVMs are compared with the classical approach adopted for hyperspectral data classification that is a conventional pattern recognition system made up of a classification method combined with a feature-reduction technique. In these two experiments, they adopted the most popular multiclass strategy used for SVMs, which is the OAA strategy. Finally, the third experiment aims at analyzing and comparing the effectiveness of the different multiclass strategies described in the previous section, which is the OAA, O[illegible] Table 2.2. Continued… OAA str[illegible]	Support Vector Machines (SVMs)	Their work shows that SVMs exhibit low sensitivity to the Hughes phenomenon, resulting in an excellent approach to avoid the usually time-consuming phase required by any feature-reduction method. However, other factor that can still affect their results was not mentioned like the configuration of the system used and size of the dataset.
32	Durgesh and Bhambhu, (2009)	Data Classification Using Support Vector Machine	In their experiment, the support vectors, which are critical for classification, were obtained by learning from the training samples. In this Data Classification using Support Vector Machine they have shown the comparative results using different kernel functions for all data samples and RSES data sets. In these experiments used LIBSVM with different kernel linear , polynomial, sigmoid and RBF kernel	LIBSVM and RBF kernel	The experiment can be tested using Medium or high resolution satellite imageries to map urban-rural objects.
33	Tarun Rao	Supervised Classificati	The SVM classification method was used to build a classification model	Support Vector	SVM used can be commercial

	Rajinikanth, (2014)	on of Remote Sensed data Using Support Vector Machine.	for two datasets. The first data set is of CSV format and the second one is a raster TIFF image. Later the classification model is validated against a test data set which is a subset of the input dataset. The performance of SVM is calculated using kappa statistics and accuracy parameters and it is established that for the given data sets SVM classifies the raster image dataset with better accuracy than the CSV dataset.	Machine	package type or authors developed SVM algorithm with four main kernel types but this was not stated.
34	Xie, (2006)	Support Vector Machines for Land Use Change Modelling	The research was to develop a novel method for land use change modelling with the capacity to effectively address land use change data, which might be a mixture of continuous and categorical variables that are not normally distributed. Pattern classifiers that fit the characteristic of land use change data and with a number of attractive features, namely, SVM, are applied to model land use change in Calgary, Canada from 1985 to 2001.	A SVMs land use change modelling framework was developed using C++ programming language.	The performance of SVMs was compared with that of a well-studied land use modeling approach, namely, spatial logistic regression (SLR), the comparison showed that SVMs were superior to SLR. However, the method of error matrix and kappa testing was not mentioned.
35	Shi and Yang, (2012)	Support Vector Machines for Landscape Mapping from Remote Sensor Imagery	They have reviewed the research status of SVM for landscape mapping with special attention on heterogeneous landscape types. Then, they have implemented this technique to map various land cover types in an urban area from a remote sensor image. Their studies further confirm that SVM can significantly outperform the ML, the most widely used pattern recognition method in the remote sensing community. They found that SVM can significantly improve mapping accuracy, particularly for spectrally and spatially complex landscape	SVM and ML	The study fails to indicate how the software or the classifier was developed or complied.

			categories.		
36	Bekkari, et al (2012)	SVM classification of high resolution urban satellites Images using Haralick features	The SVM classification was conducted using a combination of multi-spectral features and Haralick texture features as data source. They have used homogeneity, contrast, correlation, entropy and local homogeneity, which were the best texture features to improve the classification algorithm. The result was compared with both a standard SVM classifier and a SVM classifier with a Graph Cuts approach that introduces spatial domain information applied as a post-classification.	SVM using Haralick features	The workflow of this study can be used in other remote sensing application, especially, in rural areas for thematic land cover. However they did not study the different kernel choice in order to determine the appropriate one, for this type of image classification.
37	Zhang, et al (2013)	SVM-Based Classification of Segmented Airborne LiDAR Point Clouds in Urban Areas	They proposed an object-based classification method for classifying the airborne LiDAR point clouds in urban areas. In the process of classification, the surface growing algorithm was employed to make clustering of the point clouds without outliers, thirteen features of the geometry, radiometry, topology and echo characteristics are calculated, a SVM was utilized to classify the segments, and connected component analysis for 3D point clouds was proposed to optimize the original classification results. Table 2.2. Continued…	SVM-Based Classification	The experimental results suggest that their method is capable of making a good coarse urban scene classification. Thus, the method can be improve by using surface growing algorithm that will employed supervise method rather than clustering of the point clouds.
38	Bellman and Shortis (2014)	Using Support Vector Machines For Building Recognition In Aerial	A set of classification test data, in the form of square image patches, was extracted from three frames of colour aerial photography. The photographs were originally acquired for a project over the city of Ballarat in Victoria. The classification test was based on a balanced test set of 100 building	Support Vector Machines	The small size of the training set should be increased in order to boost the generalization of broader

		Photographs	images and 100 non-building images. Image coefficients were extracted using the wavelet process described earlier. A public domain Support Vector Machine (Joachims 1998) was used to classify the image patches into building or non-building categories. The image patches used to train the SVM classifier using several different kernels including polynomial and sigmoidal kernels.		dataset.
SECTION V *SVM Application in Africa and Nigeria in General*					
39	Kwesi, (2012)	Oil Palm Mapping Using Support Vector Machine With Landsat ETM+ Data	Nooni, mapped oil palm related land cover of a section from northern portion of Ejisu- Juaben district in the Ashanti Region of Ghana using SVM with Landsat ETM+. The Landsat ETM+ data acquired in 2010 was used for processing and image classification. Field data were acquired through stratified random sampling. A total of 343 samples were collected for classification and accuracy assessment. The classification was carried out using ML and SVM based on best three band combination from the image. The SVM and ML performance evaluation was done using overall accuracy assessment and kappa statistics procedure. The results of separability analysis showed that ETM+ data provides spectral Table 2.2. Continued… ypes	MLC and SVM	Mapping using HRSI is required because the result from and kappa statistics in both classifier were below 0.80.
40	Owamoyo, et al (2013)	Number Plate Recognition For Nigerian Vehicles	They present the number plate extraction, character segmentation and recognition work, with English characters. Number plate extraction was done using Sobel filter, morphological operations and connected component analysis. Character segmentation is done by using connected component and vertical projection analysis. Character recognition is carried out using SVM. Experiments show that the algorithm has good performance	SVM using MATLAB.	This work was implemented using MATLAB. The reason (s) of using Sobel filter andSVM was not stated and it could have been better to use one of algorithm for

			on number plate extraction, and character segmentation work. It can deal the images correctly, with noise, illumination variance, and rotation to ±50. This work is implemented using MATLAB.		both Number plate extraction andCharacter recognition while the other may be used for comparison of their performance.
41	Okwuashi, et al (2012).	Predicting Future Land Use Change Using Support Vector Machine Based GIS Cellular Automata: A Case Of Lagos, Nigeria	This research used the SVM based GIS CA calibration for land use change prediction of Lagos. The SVM based CA model is loosely coupled with the geographic information systems. SVM parameters are optimised with the k-fold cross-validation technique, using the linear, polynomial, and RBF kernels functions. The land use change prediction is based on three land use epochs: 1963-1978, 1978-1984, and 1984-2000. The performance of the model was evaluated using the Kappa statistic and receiver operating characteristic and The modelling was implemented in MATLAB and visualised in ArcGIS. The training data were extracted using the stratified random sampling. The training data consist of developed and undeveloped cells. Developed cells were labelled +1 while undeveloped cells were labelled -1.	Support vector machine based GIS cellular automata and the modelling was implemented in MATLAB and visualised in ArcGIS.	Their stratified random sampling of training data and theevaluation of performance of the model using the Kappa statisticin MATLAB required further studies.
42	Ndehedehe, (2014)	Support Vector Machine Based Kernel Types in Extraction of Urban Areas in Uyo Metropolis from Remote Sensing Multispectral Image	Applied a pixel based classification to classify built up areas using a carefully extracted training sample as a spectral signature of the specified region of interest. The SVM kernel type implementation was software-based and the results were validated using Orthophoto-derived vector of the study area. Then, the classified data using various SVM kernel types were vectorised and compared with the existing vector map of the same location using a GIS approach. The class statistics of the kernel methods used was extracted from ENVI 4.7 after the classification and the post	Support Vector Machine Based Kernel Types in SVM SVM in ENVI ver.4.7 commercial software	This study used a carefully extracted training sample as a spectral signature of the specified region of interest. But the training sample separabilities index analysis was not done or tested.

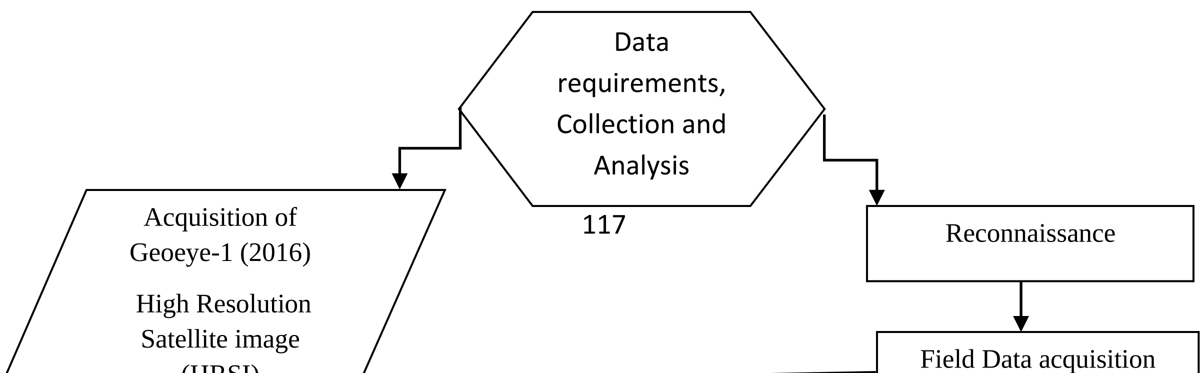
Table 2.2. Continued…

			classification results of the built up pixels from the 4 kernel obtained.		
43	United Nations Human Settlements Programme (UN-HABITAT), 2009	Executive Summary of Structure Plans for Awka, Onitsha and Nnewi2027	The method adopted was the participatory and very engaging Rapid Urban Sector Profiling for Sustainability (RUSPS), first developed by the UN-HABITAT. and has been successfully employed in several countries. The UN-HABITAT organized a week-long training on RUSPS to properly induct 10 Nigerian Experts and 15 governments technical staff drawn from Anambra State Ministries and local governments and Federal Ministry of Housing and Urban Development in RUSPS methodology. During the training, RUSPS global data collection framework was reviewed with an agreement on the under-mentioned seven thematic issues for the conduct of RUSPS for Anambra cities: Basic urban services; Environment, Gender, Governance, Heritage/historic areas, Local Economic Development; and Shelter and slums.	IKONOS Satellite image, Personal observation, statistical records and ArcGIS software	The method used in producing their landuse and landcover maps was not recorded, the statistical records used is not enough for this kind of study but required a well-structured questionnaire and the year their IKONOS Satellite image was acquired was not revealed in their report.

CHAPTER THREE

METHODOLOGY

This chapter describes the methodology that was used in carrying out the study,and discussed under Data requirement, hardware and software requirement, Data acquisition, modelling and processing. The research methodology was based on remotely sensed data processing and SVMs application to produce and visualize the land information of the area of study.The flow chart of the methodology (Fig.3.1) focuses on the data needs and the methods used in achieving the set aim and objectives of this study.



Note: RO/RQ means Research Objective and Research Question 1-5

Figure 3.1: Flowchart of Methodology

3.1 Data requirements

The Hardware and software in section 3.2 were employed in the analysis and production of this project.

3.2 Hardware and Software Data Requirement

3.2.1 Hardware requirement

The Hardware that was used in this project was basically for data transfer, processing, analysis and graphic plotting. The following hardware was used:

i. Garmin ‘GPSmap 78sc’ Global Positioning System (GPS)

ii. Digital camera

iii. External hard drive

iv. A personal Hp Elitebook workstation Laptop Computer Windows 10 Professional (2005) with the following configurations; 780GB Hard disk, 4GB RAM, 4.0GHZ, Intel Processor Core i7 and 64 bit Operating System and 2GB Dedicated.

v. A personal Zinox Laptop Computer Windows 7 Professional (2009) with the following configurations; 320GB Hard disk, 4GB RAM, 4.0GHZ, Intel Processor Core i7 and 64 bit Operating System

vi. HP Deskjet 1102 (A4) Black & White Printer

vii. HP LaserJet K7100 series A3 Colour Printer

3.2.2 Software used

The software that was used for this project includes:

i. ENVI (Environment for visualizing images) software version 5.1

ii. ILWIS (Integrated Land and Water Information System) Academic 3.3 Version

iii. The ArcGIS 10.1 software

iv. GraphPad Prism, version 7.1

v. Statistical Package for Social Sciences (SPSS Version 21.0)

vi. Microsoft word, Excel and Power Point 2010

3.3 Data Acquisition, Modeling and Processing

3.3.1 Data sources and acquisition

Data collections include the following:

i. Administrative map of Nigeriawas obtained from office of the Surveyor General of the Federation (SGoF) and modified in the Remote Sensing and GIS Laboratory, Department of Environment Management, Chukwuemeka Odumegwu Ojukwu University (COOU), (2017).

ii. Administrative map of Anambra State and Nnewi North LGA was obtained from office of the Surveyor General of the Federation (SGoF) and modified in the Remote Sensing and GIS Laboratory, Department of Environment Management, Chukwuemeka Odumegwu Ojukwu University, (2017)

iii. Topographical map of study area (published in 1964) was obtained from the Ministry of Surveys and Town Planning, Awka.

iv. High Resolution Satellite image of Nnewi North LGA (GeoEye-1 Image of 2016) was obtained from DigitalGlobe online in ECW format (Enhanced Compressed Wavelet; a proprietary data format developed by Earth Resource Mapping that is primarily intended for aerial imagery) consequently, transform to ENVI.dat format and GeoTagged Image File Format (GeoTIFF), SVM Algorithm is a classifier used for image classification. Algorithm is a step-by-step procedure for performing a specific task, such as a mathematical formula or a set of instructions on a computer program.(The data was stored in the internal and external hard Drive).ENVI (Environment for visualizing images) software version 5.1 was the main software that houses the SVM algorithm used for the analysis.

GeoEye Inc. (formerly Orbital Imaging Corporation or ORBIMAGE) was an American commercial satellite imagery company based in Herndon, Virginia. GeoEye was merged into the DigitalGlobe Corporation January 29th, 2013. (Digitalglobe, 2015). The company was founded in 1992 as a division of Orbital Sciences Corporation in the wake of the 1992 Land Remote Sensing Policy Act which permitted private companies to enter the satellite imaging business. The division was spun off in 1997. It changed its name to GeoEye in 2006 after acquiring Denver, Colorado based Space Imaging for $58 million, (John Pike 2006). Space Imaging was founded and controlled by Raytheon and Lockheed Martin. Its principal asset was the IKONOS satellite. GeoEye-1, is one of the world’s highest-resolution commercial colour imaging satellite, was launched on September 6, 2008 from Vandenburg Air Force Base in California. The satellite offers extraordinary detail, high accuracy and enhanced stereo for DEM generation. GeoEye-1 will simultaneously collect panchromatic imagery at 0.41m and Multispectral imagery at 1.65m. Due to U.S. Government Licensing, the imagery will be made available commercially as 0.5m imagery. GeoEye-1 satellite has the capacity to collect up to 700,000 square kilometers of panchromatic imagery (and up to 350,000 square kilometers of Pan-Sharpened Multispectral imagery) per day. The GeoEye-1 Satellite Sensor Specifications and its technical characteristics are given in table 3.1 and while the image of the sensor is displayed as Fig 3.2 also the properties of subset raster dataset (GeoEye-1) Satellite image used are shown in table 3.2.

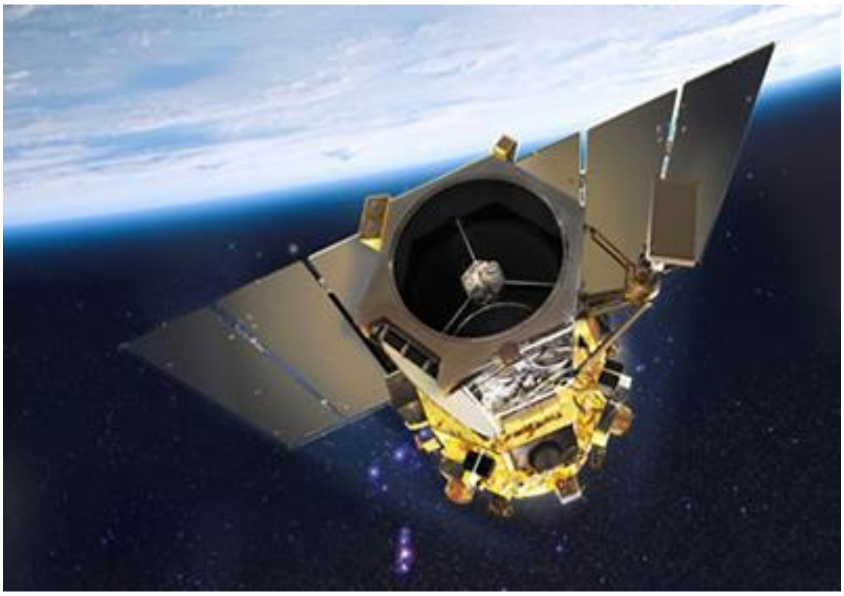


Figure 3.2: GeoEye-1 satellite image Sensor

Source: info@geoeye.com (2016) GeoEye-1 Fact Sheet, http://www.geoeye.com.

Table 3.1: The GeoEye-1 Satellite Sensor Specifications and its technical characteristics

GeoEye-1 Satellite Sensor /image Specifications

Launch Date	September 6, 2008
Camera Modes	Simultaneous panchromatic and multispectral (pan-sharpened) Panchromatic only Multispectral only
Resolution	0.46 m / 1.51 ft* panchromatic (nominal at Nadir) 1.84 m / 6.04 ft* multispectral (nominal at Nadir)
Spectral Range	Panchromatic: 450 - 800 nm Blue: 450 - 510 nm Green: 510 - 580 nm Red: 655 - 690 nm Near Infra-Red: 780 - 920 nm
Metric Accuracy/Geolocation	5 m CE90, 3 m CE90 (measured)
Swath Widths & Representative Area Sizes	Nominal swath width - 15.2 km / 9.44 mi at Nadir Single-point scene - 225 sq km (15x15 km) Contiguous large area - 15,000 sq km (300x50 km) Contiguous 1° cell size areas - 10,000 sq km (100x100 km) Contiguous stereo area - 6,270 sq km (224x28 km) (Area assumes pan mode at highest line rate)
Imaging Angle	Capable of imaging in any direction

Revisit Frequency at 770 km Altitude (40° Latitude Target)

Max Pan GSD (m)	Off Nadir Look Angle (deg)	Average Revisit (days)
0.42	10	8.3
0.50	28	2.8
0.59	35	2.1

Daily Monoscopic Area Collection Capacity	Up to 700,000 sq km/day (270,271 sq mi/day) of pan area (about the size of Texas). Up to 350,000 sq km/day (135,135 sq mi/day) of pan-sharpened multispectral area (about the size of New Mexico)

TECHNICAL INFORMATION

Launch Vehicle	Delta II
Launch Vehicle Manufacturer	Boeing Corporation
Launch Location	Vandenberg Air Force Base, California
Satellite Weight	1955 kg / 4310 lbs
Satellite Storage and Downlink	1 Terabit recorder; X-band downlink (at 740 mb/sec or

	150 mb/sec)
Operational Life	Fully redundant 7+ year design life; fuel for 15 years
Satellite Modes of Operation	Store and forward Real-time image and downlink Direct uplink with real-time downlink
Orbital Altitude	770 km / 478 miles
Orbital Velocity	About 7.5 km/sec or 17,000 mi/hr
Inclination/Equator Crossing Time	98 degrees / 10:30am
Orbit type/period	Sun-synchronous / 98 minutes

Source: https://www.satimagingcorp.com/satellite-sensors/geoeye-1/

Table 3.2: Properties of Raster Dataset (GeoEye-1) Satellite image used

Raster Properties	
----Extents Properties ----	**---Coordinate System Properties ----**
Extent Left = 6.87246304 Extent Top = 6.07371183 Extent Right = 6.96037837 Extent Bottom = 5.965178 Units = Degrees, Columns= 16480, Rows = 20345, Number of Bands = 3 Size = 4,023,427,200 bytes (Unmasked)) Size = 335,285,600 bytes(Masked) Map Tie Point X = 6.87246304 Map Tie Point Y = 6.07371183 Pixel Size X = 0.00000533 Pixel Size Y = 0.00000533	GEOGCS = "GCS_WGS_1984" Datum = "D_WGS_1984" Spheroid = "WGS_1984",6378137.0,298.257223563 Prime Meridian = "Greenwich",0.0 Unit = "Degree",0.0174532925199433,Acquisiti on Time = 1/24/2016.

Table3.2is showing the raster properties, map info properties, coordinate system properties, extents properties, spectral properties, time properties among others of image used.

3.3.2 Data modeling and processing

The study used ENVI software version 5.1 was used for image processing, image enhancement, Filtering and masking and Image classification. Meanwhile, ArcGIS 10.1 was employed for producing thematic maps. Microsoft Excel, GraphPad Prism and SPSS were employed for data analysis and results presentation.

3.3.2.1 *Data pre-processing*

The initial preparation for the processing of any remotely sensed image data is done under pre-processing. This involves reading the user requirements and the input data and storing these information & data for further processing in the appropriate format under a project area. The input data was radiometrically and geometrically corrected by GeoEye image cooperation. Data georeferencing and transformation was also carried by vendors. What is transformation? Transformation is an image processing operation that changes data to another data space, usually by applying a linear function. The goal of most transformations is to improve the presentation of information. Transformed images are often more easily interpreted than the original data.

3.3.2.2 *Data processing*

The following processing procedures were carried out:

i. Image enhancement
ii. Image Sub-setting
iii. Extraction of Features of Interest(FOIs)
iv. Supervised classification using SVM, and ML
v. Post-processing Accuracy Assessment
vi. Statistical Analyses
vii. Preparation of maps

3.3.2.3 *Adoption of landuse/ landcover feature types*

For the purpose of this study and based on the knowledge of the study area, the following five class categories in table 3.3 were adopted.

Table 3.3: Adopted landuse/landcover features

S/N	Code/Class	Description	Colour assigned
1	#1Built-up	Includes all residential, commercial and industrial development	Mars Red
2.	#2Farm land	Includes, gardens, field crops, Horticulture, orchards, improve pasture, ploughed fields and fallow land.	Solar Yellow
3.	#3Vegetation	Includes all vegetation features such as evergreen, deciduous, scrubs and forest.	Leaf Green
4.	#4Water bodies	Water related features such as freshwater, lakes, rivers and streams.	Lapis Lazuli
5.	#5Open/bare surface	Includes bare earth or soil, unpaved roads, erosion sites,playground and excavation sites.	Indicolite Green

3.3.3 Procedures adopted in achieving objective number one.

Objective No. 1"To identify and extract features of interest (FOIs) in a subset of HRSI of the study area" was achieved as follows:

Supervised classification requires that the operator be familiar with the area of interest in the area covered by the image. The operator needs to know where to find the classes of interest in the area covered by the image; this information can be derived from general area knowledge or from dedicated field observations. Natural and Man-Made Objects which include the vegetation, wetlands/water bodies, buildings/pavement, open/bare surfaces and farm lands were captured in High Resolution Satellite Image (HRSI) and classified using Support Vector Machine (SVM) Classifier with a view of developing an up-to-date and reliable urban land cover map. This was carried out after image enhancement and Sub-setting also known as chipping in the ENVI environment.

3.3.3.1. *Extraction offeatures of interests(FOIs)*

The number one objective which is “Extraction of Features of Interests (FOIs)” was careful carried out bycreating FOIs from geometry. First is to download and save the image data in a computer, link the image to a folder and launch the ENVI. To process images using ENVI+IDL starts by launching the ENVI Version 5.1, incorporated with Interactive Data Language 8.3 (IDL). IDL is a high-level programming language that contains an extensive library of image processing and analysis routines. With IDL, you can quickly visualized image data and begin investigating the best way to extract useful information.

A digital image is composed of a grid of pixels and stored as an array. A single pixel represents a value of either light intensity or colour. Images are processed to obtain information beyond what is apparent given the image’s initial pixel values. Image processing tasks carried out include the combination of the following: Modifying the Image View: Transforming, translating, rotating and resizing images are common tasks used to focus the viewer’s attention on a specific area of the image. Enhancing Contrast and Filtering: Contrasting and filtering provide the ability to smooth, sharpen, enhance edges and reduce noise within images.

3.3.4. Procedures adopted in achieving objective number two.

Objective No. 1“To evaluate the result of region of interests and ground truth extracted using the Jeffries-Matusita and Transformed Divergence separability index” was achieved as follows:

3.3.4.1. FOI separability index analysis

Features of Interest (FOIs) known as Region of Interests (ROIs) inENVI software are selected samples of a raster, such as areas of water, farm land, vegetation, open surface which are identified for a particular purpose of image classification. Modern software for satellite image processing offers its users a wide range of supervised classification algorithms. It yields powerful capabilities for automation of the image interpretation process. In return for that, a user should make training areas of high quality. It is this quality that defines the accuracy of the supervised classification.ENVI allows you to analyze panchromatic, multispectral, hyperspectral, radar images, and also works with digital elevation models data. However, the highest work in image classification is the training of good FOIs. This research uses the**n-D Visualizer** to locate, identify, and cluster the purest pixels and the most extreme spectral responses (endmembers) in a dataset in n-dimensional space.

You typically use the n-D Visualizer with spatially subsetted Minimum Noise Fraction (MNF) data that use only the purest pixels determined from the Pixel Purity Index (PPI).You can export selected FOIs to the n-D Visualizer so you can see the distribution of the points within and between the FOIs. This option is useful for checking the separability of your classes when you use FOIs as input into supervised classifications which is also known as Bhattacharyya distance analysis.

For good classification results using these FOIs, the groups of pixels for the different FOIs should be separate from each other and should not overlap. If the pixels' overlap, edit the groups of pixels by selecting the appropriate colours from the Class menu to add pixels to a FOI or by selecting White to remove pixels from a FOI. The research used the **ROI Separability** tool to compute the spectral separability between selected ROI pairs for a given input file. Both the Jeffries-Matusita and Transformed Divergence separability measures are reported. These values range from 0 to 2.0 and indicate how well the selected FOI pairs are statistically separated. Values greater than 1.9 indicate that the FOI pairs have good separability. For FOI pairs with lower separability values, you should attempt to improve the separability by editing the ROIs or by selecting new ROIs. For FOI pairs with very low separability values (less than 1), you might want to combine them into a single ROI.

3.3.5. Procedures adopted in achieving objective number three.

Objective No. 3 "To perform image classification using Support Vector Machine in ENVI 5.1 Software" was achieved as follows:

i. Launched the software and input the raw satellite image in the system (Fig. 3.5)

ii. Design image classification scheme by extracting ROIs: ROIs are usually information classes such as Built-up area, Farm lands, vegetation areas, water bodies, open/bare surface among other classes of interest. Conduct field surveys and collect ground information and other ancillary data from the study area.

iii. Pre-processing of the image, including radiometric, atmospheric, geometric and topographic corrections, image enhancement,Georeference (if not georefenced), image subsetting(Fig. 3.6) and initial image clustering.

iv. Select representative areas of the image, "Features of Interests" (FOIs) and analyze the initial clustering results or generate training signatures.

v. Run Image classification algorithms of your choice.

vi. Post-processing: complete geometric correction & filtering and classification decorating.

vii. Accuracy assessment: compare classification results with field studies. (Gong and Howarth 1990) and

viii. Production of maps and results from the classification image

3.3.5.1. *Image classification techniques*

There are various classification approaches that have been developed and widely used to produce land use and land cover maps. They include supervised to unsupervised; parametric to non-parametric to non-metric, or hard and soft (fuzzy) classification, or per-pixel, sub-pixel, and prefield among others but SVM supervisedimage classification was adopted in this research. The majority of image classification is based solely on the detection of the spectral signatures (i.e., spectral response patterns) of land cover classes.

The success with which this can be done will depend on two things:

1) The presence of distinctive signatures for the land cover features of interest in the band set being used; and

2) The ability to reliably distinguish these signatures from other spectral response patterns that may be present.

There are two general approaches to image classification: supervised and unsupervised. They differ in how the classification is performed. In the case of supervised classification, the software system delineates specific landcover types based on statistical characterization data drawn from known examples in the image (known as Region of Interests (ROIs) in ENVI). With unsupervised classification, however, clustering software is used to uncover the commonly occurring landcover types, with the analyst providing interpretations of those cover types at a later stage.

3.3.5.2. ***Supervised classification***

The first step in supervised classification is to identify information on features of interest in the image, called regions of interest in ENVI or training sites in other remote sensing softwares. The software system is then used to develop a statistical characterization of the reflectances for each information class. This stage is often called signature analysis and may involve developing a characterization as simple as the mean or the range of reflectances on each band, or as complex as detailed analyses of the mean, variances and covariances over all bands.Once a statistical characterization has been achieved for each information class, the image is then classified by examining the reflectances for each pixel and making a decision about which of the signatures it resembles most. There are several techniques for making these decisions, called classifiers.

The supervised image classification was carried using SVM and ML. The general procedure is as follows:

(i) Input the classifying image and associate it to already trained ROIs or you can train it straight away.

(ii) Wait for a while for the system to automatically compute the decision rule/classification statistics of all the selected ROIs

(iii) Classify the image according to the classifier selected

(iv) The result automatically displays on the screen.

However, it is important to note that you have to wait patiently for 30-45 minutes for the image to be classified using ML. For SVM, you will have to wait for about 72-96 hours before the classification will be done and displayed on the screen of the computer. From here you can save the result to any format of your choice e.g. ENV.tiff. After the image classification in ENVI the results were converted from raster to vector automatically by clicking at the icon "conversion to vector".

3.3.6. Procedures adopted in achieving objective number four.

Objective No. 4 "To evaluate the performance of SVM and ML in mapping geometric features using Error Matrix, Kappa and a newly proposed Post-Confusion Matrix (PoCoMa) template" was achieved as follows:

3.3.6.1. *Calculation of confusion matrices*

Confusion matrix can be used to show the accuracy of a classification result by comparing a classification result with ground truth information. ENVI can calculate a confusion matrix (contingency matrix) using either a ground truth image or using ground truth ROIs to assess classification accuracy and misclassification between categories automatically. In each case, errors of commission and omission are computed in order to obtain their overall, producer, user and kappa coefficient accuracies. The matrix is size m x m, where m is the number of classes. The rows in the matrix represent classes that are assumed to be true, while the columns represent classes derived from remote sensing imagery.

Several sites were visited (ground truthing) in the study area with respect to the adopted classes of land use and land cover. Coordinates of the sites were also taken. Reference datacollected through the process of "ground truthing", determination of class types at specific locations and Compare reference to classified map. Having done that you have to ask yourself, does class type on classified map equal to class type determined from reference data? If yes, it is ok and if not, then the analyst has to go back to drawing board because the goals of accuracy assessment are to assess how well a classification worked and to understand how to interpret the usefulness of someone else's classification.

3.3.6.2. ***Calculation of kappa coefficient***

The kappa coefficient determines the difference between the observed agreement between two maps and the agreement that might be attained by change matching of the two maps. It gives a measure of agreement, Table 3.4.

The Kappa coefficient is given in Pradhan *et al.*, (2013) as:

$$K = \frac{N \sum_{i=1}^{r} X\text{ii} - \sum_{i=1}^{r} (Xi_{+} \times X + 1)}{N2 - \sum_{i=1}^{r} (Xi_{+} \times X + 1)}, \qquad 5.1$$

Where

r- Number of rows in the confusion matrix

N- Total number of observations included in matrix

Σx_{ij} - Number of observations along the major diagonal

(X_{1+})- Total observations in row i shown as marginal total to the right of the matrix

(X_{+1})- Total observation in column i shown as marginal total at bottom of the matrix r

The kappa statistic is more reliable than other validation techniques because it has the ability to evaluate the actual agreement and chance agreement (Fung and Ledrew, 1988).

Table 3.4: Interpretation of kappa statistic

Kappa	Interpretation
< 0	No agreement
0.0-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-1.00	Almost perfect agreement

Source: Fung and Ledrew,(1988).

3.3.6.3.*Post confusion matrix statistical analyses*

The studywas not satisfactory analyzing the performance of SVM and ML in mapping geometric features using only overall accuracyand Kappa Statistical Procedure, thus,developed a “Post Confusion Matrix” (PoCoMa) formula from the contingency matrix, kappa coefficient and the behaviours of the image classifiers used, for computation and comparison of their capabilities. The parameters that were designed and used are in table 3.5.

Table 3.5: Post Confusion Matrix (PoCoMa)Parameters

S/N	PARAMETERS CODE	DESCRIPTIONS
1	OA	Overall Accuracy

2	KC	Kappa Coefficient
3	UA	User Accuracy
4	PA	Producer's Accuracy (Average Value)
5	EOO	Error of Omission (Average Value)
6	EOC	Error of Commission (Average Value)
7	SPD	Speed/Time (Divided By 3600)
8	SP	Space/Memory
9	RB	Rule Based
10	SV	Supervised
11	ROI	Features of InterestROIs
12	COI	Classes of Interest (Average Value)

The values obtained were tested using t-statistics (Paired Samples Test), where GraphPad Prism version 7.0 and SPSS version 21 statistical software were employed for the data analyses.

3.3.6.4. ***The Null Postulated statement was tested***

(1) Hypotheses one

Null hypothesis (H_0): There is no significant difference between the performance of SVM and MLC.

3.3.7. Procedures adopted in achieving objective number five.

Objective No. 5 "To produce the land use and land cover map of Nnewi-North L.G.A", that will serve as a base map for improved land-use planning and monitoringby end-users

The landuse and landcover map that will serve as a base map of Nnewi North was produced using ENVI SVM and ML algorithms. After Image classification, post-processing, (including ground truthing/field verification) and computation of the error matrix to assess accuracy of the work, then the result of image classification will be accepted if satisfactory, otherwise you may reject the result and retrain the image for better performance. When this necessary activity has been achieved, the map will be designed cartographically showing the necessary features of interest, thereafter, the result will be publishingin a digital format or print as a hardcopy as the case may be. So, at this stage, the result of classified image in ENVI software wasexported to ArcGIS 10.2 as vector layer for cartographical analysis and design. Here, the north direction, legend, scale (graphical and absolute statement scale), index, text, border line, grid lines and coordinates, title of the map among others were added to the data frame of the classified dataset to cartographically modified the proposed output map using the mentioned conventional symbols that can be easily understood globally by professional and unskilled people alike.

CHAPTER FOUR

RESULTS AND DISCUSSION

This chapter focuses on the data analysis, the discussion and presentation of the various results obtained for this study. The results of the study were presented in form of maps,profile, figures, platesand graphs.

4.1. Procedures Adopted in Analysing Objective Number One.

Objective No. (1) "To identify and extract Features of Interest(FOIs) in a subset of HRSI of the study area"was analysed as follows:

4.1.1. Getting the relevant hardware and software

First thing is to download and save the image data in a computer, link the image to a folder and Launch the ENVI. To process image using ENVI+IDL (Environment for visualizing images+Interactive Data Language) it starts by launching the ENVI Version 5.1, IDL is a high-level programming language that contains an extensive library of image processing and analysis routines. With IDL, you can quickly visualize image data and begin investigating the best way to extract useful information, see figure 4.1, 4.2, 4.3 and 4.4.

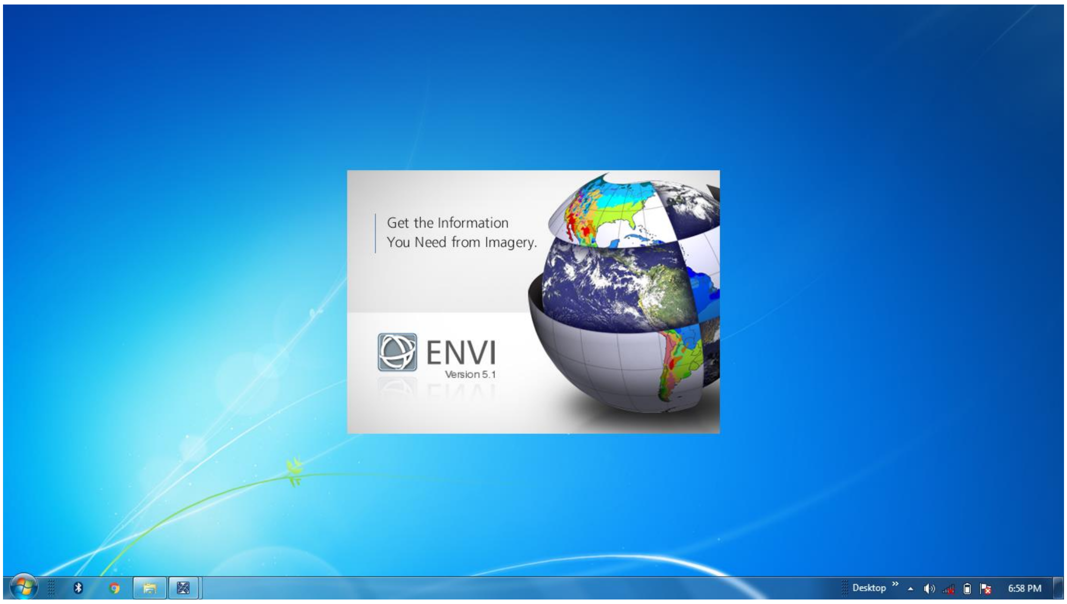


Figure 4.1: Showing ENVI 5.1 window + IDL 8.3 algorithm after it was launched

Source & Specifications: Exelis Visual Information Solutions, Inc(2013). ENVI 5.1Software +IDL Version 8.3.

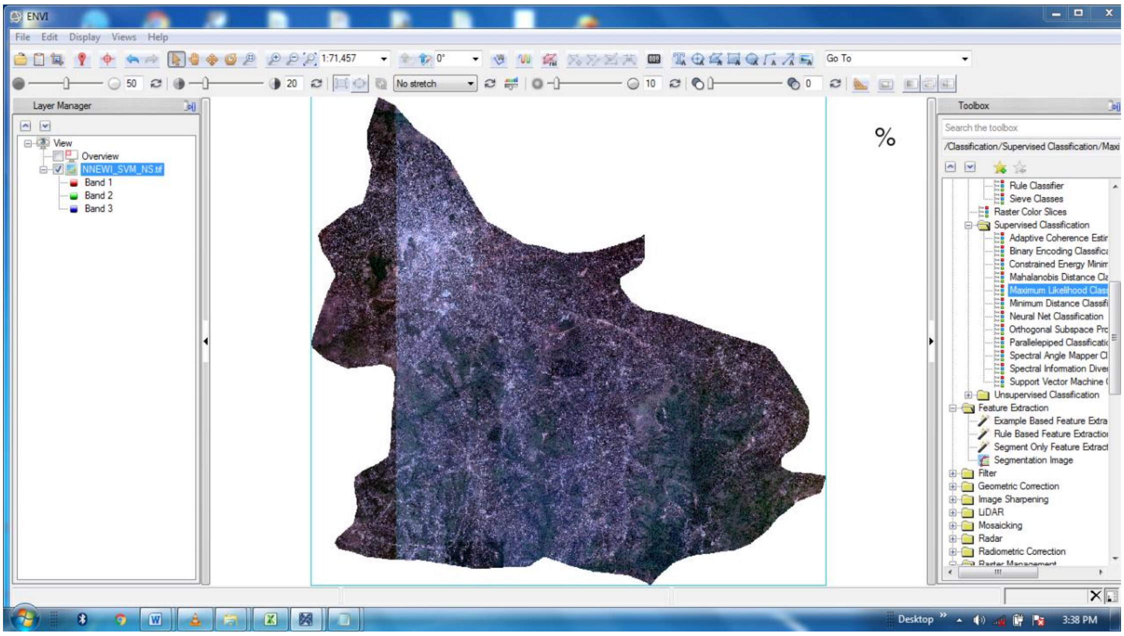


Figure 4.2: TheGeoEye-1(2016),RGB Raw Satellite Image covering Nnewi North and South Local Government was downloaded and enhanced.

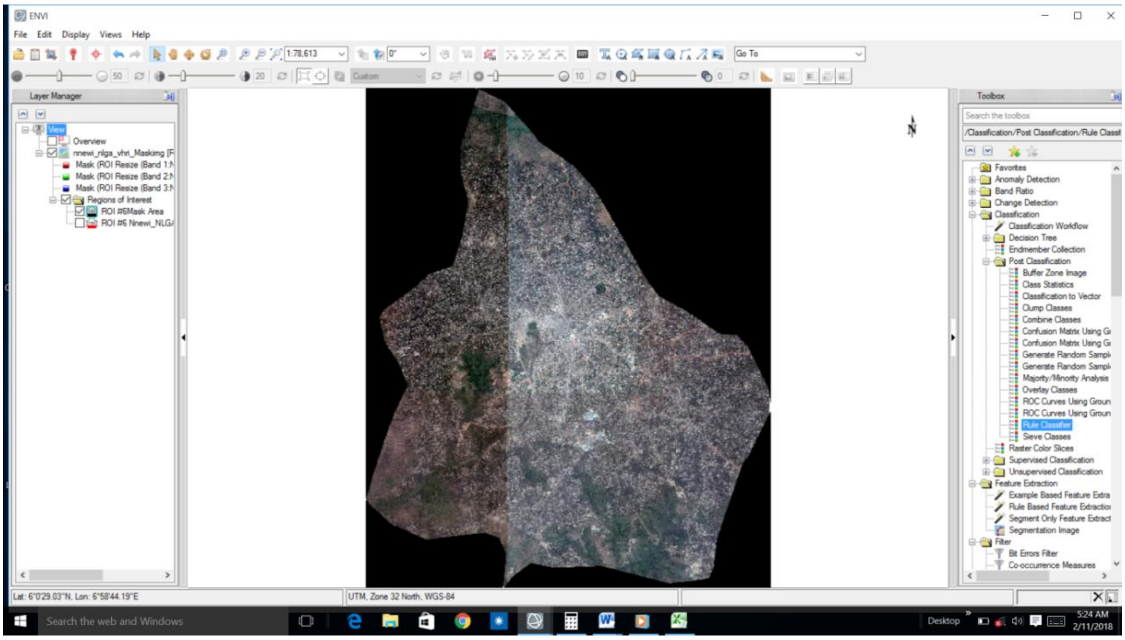


Figure 4.3: The subset of GeoEye-1 Satellite Image covering Nnewi North Local Government Area.

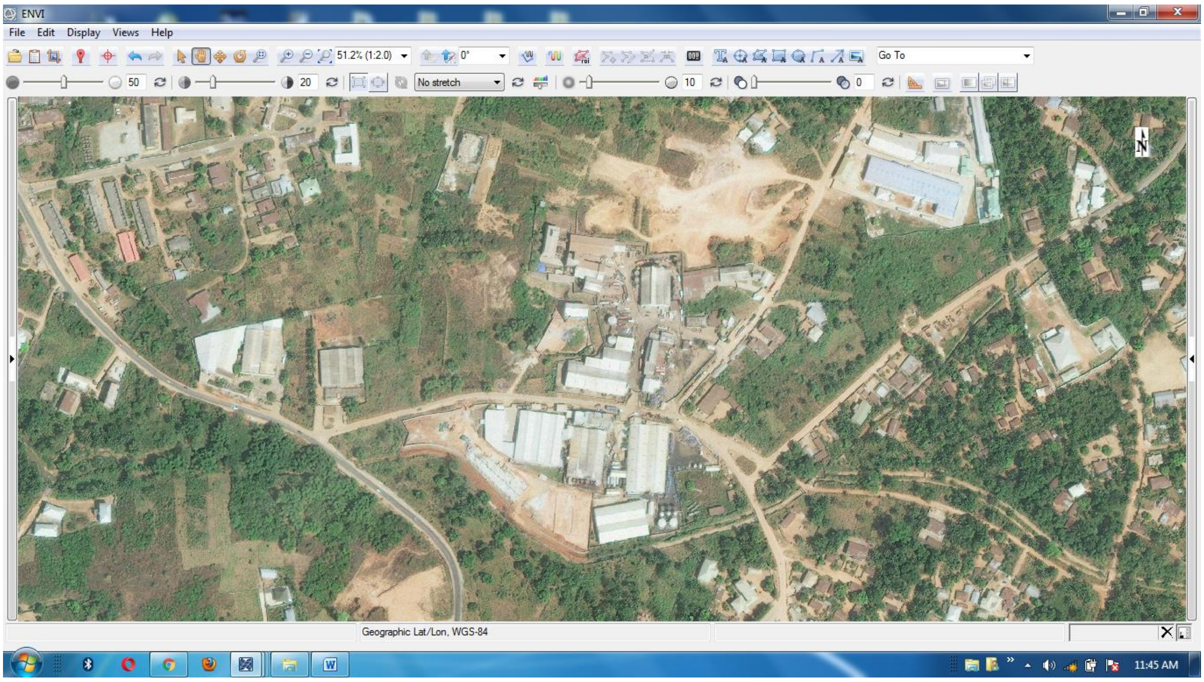


Figure4.4: Sample ofvisualized raster image of Umudim Nnewi Industrial zone in ENVI 5.1 window + IDL 8.3 system.

4.1.2. Extraction of features of interests (FOIs)

To create FOIs from geometry, draw shapes or mark points on the raster layer. You can also create multi-part FOIs, which are polygon, rectangle, or ellipse shapes that contain a hole.

What is the significance of the exercise? Supervised classification requires that the analyst be familiar with the area of interest in the area covered by the image. The analyst needs to know where to find the classes of interest in the area covered by the image; this information can be derived from general area knowledge or from dedicated field observations. A sample of a specific class comprising of a number of training pixels, forms a cluster in the feature space. The clusters as selected by the analyst should form a representative data set for a given class; this means that the variability of a class within the image should be taken into account. Thus, the number one objective which is “Extraction of Features of Interests (FOIs)” was careful carried and after the analysis the results were saved to be used for further analysis, see example some of Features of Interests (FOIs) in Fig.4.5 to 4.14.

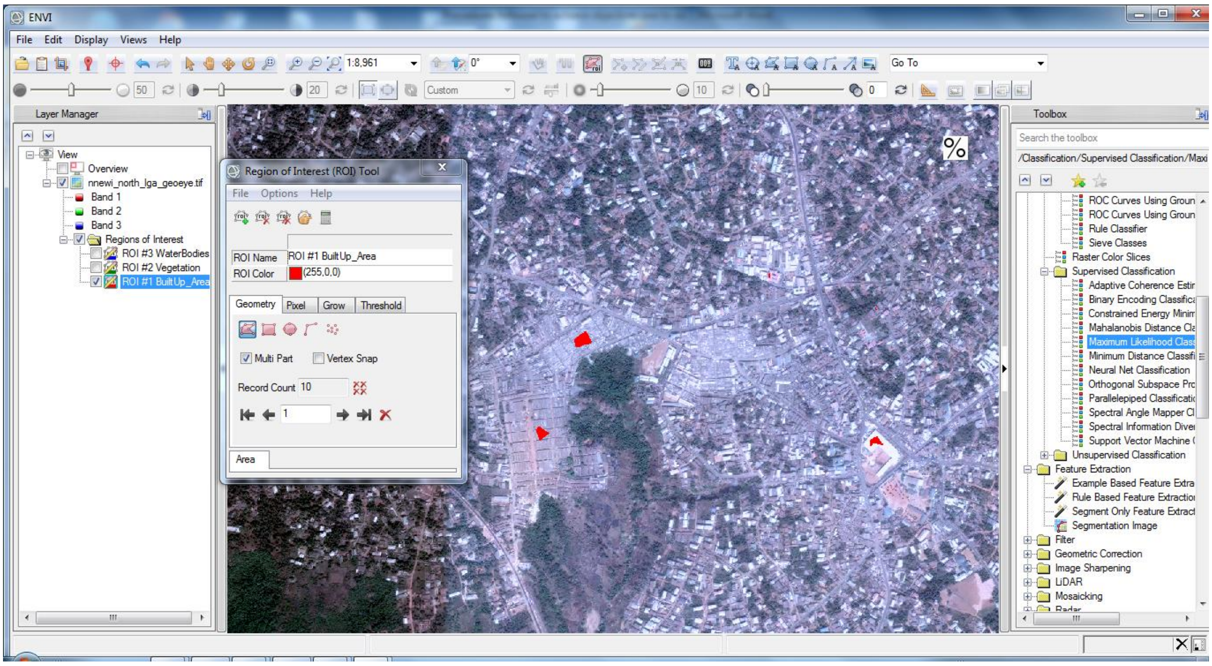


Figure 4.5: Location of built-up areaas a feature of interest at Nkwo-Nnewi

The FOI (built-up area) showing the Nkwo-Nnewi and the shopping mall under construction Fig. 4.5, in order words the known Central Business District (CBD) of Nnewi.

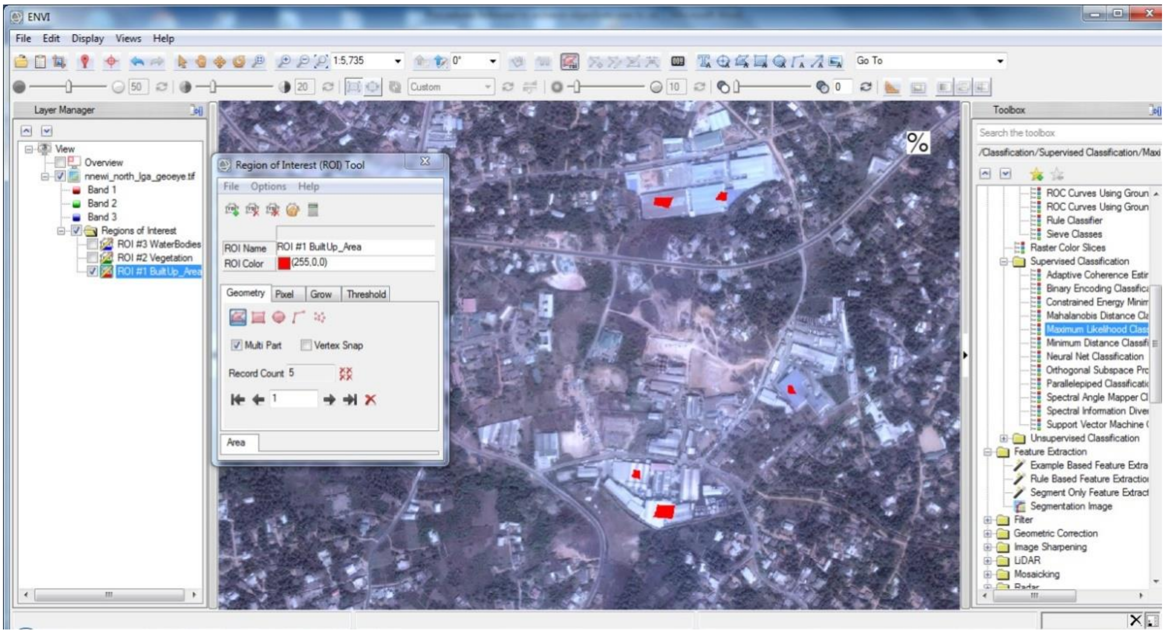


Figure 4.6: Location of Built-up as a feature of Interest atUmudim

FOI of Built-up Area showing Umudim-Nnewi industrial Area. This zone is the home for all big industries in Nnewi, Fig.4.6.

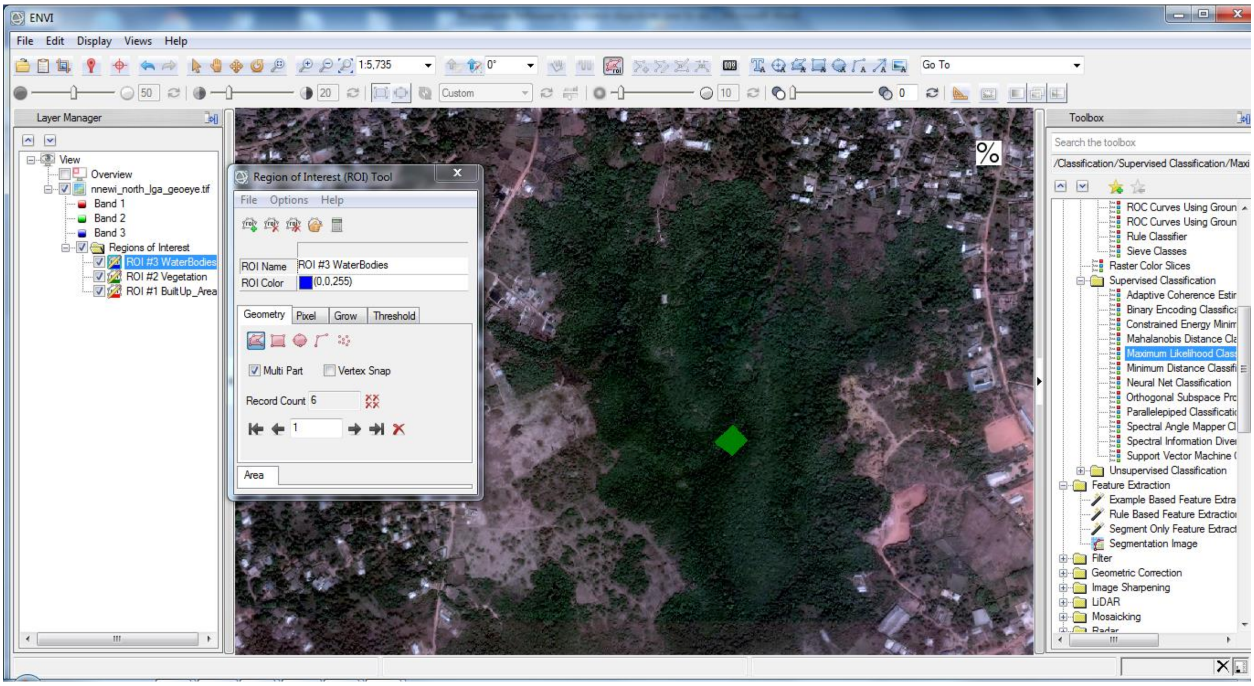


Figure 4.7: FOI of thick but secondary Vegetation at Uruagu/Umudim-Nnewi

FOI of thick but secondary Vegetation at Uruagu/Umudim-Nnewi, this area consists of valley with a secondary re-growth of vegetation and likely the source of Ele River, Fig. 4.7.

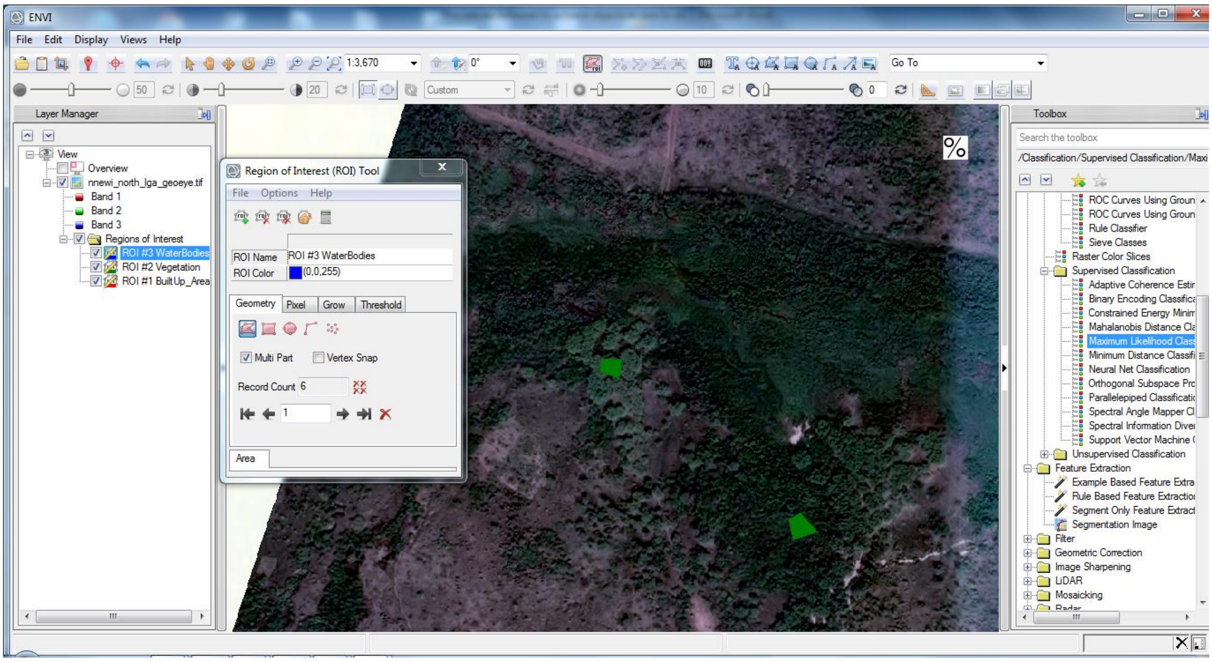


Figure 4.8: FOI of thick but secondary Vegetation at Nnewichi-Nnewi near Idemili River

FOI of thick but secondary Vegetation at Nnewichi-Nnewi near Idemili River, this area consists of a reasonable secondary with few thick kind of vegetation, Fig. 4. 8.

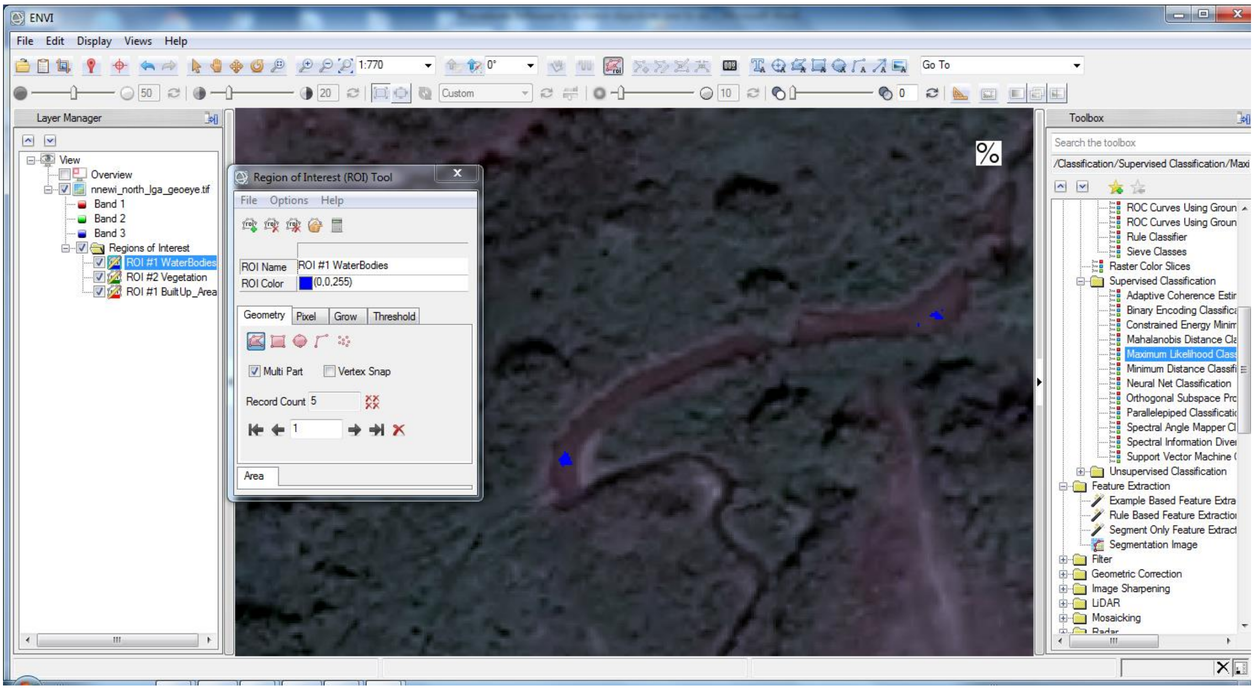


Figure 4.9: Water bodies FOI showing the ‘Mmiri-Ele’ Ele River

FOI Water bodies showing the ‘Mmiri-Ele’ Ele River, it was observed that the colour is very close to that of farm land and old houses with corroded roof,Fig. 4.9. This may as result of change of water quality caused by anthropogenic activities, coastal erosion, siltation due to seasonal flooding, and surface water pollution among others.

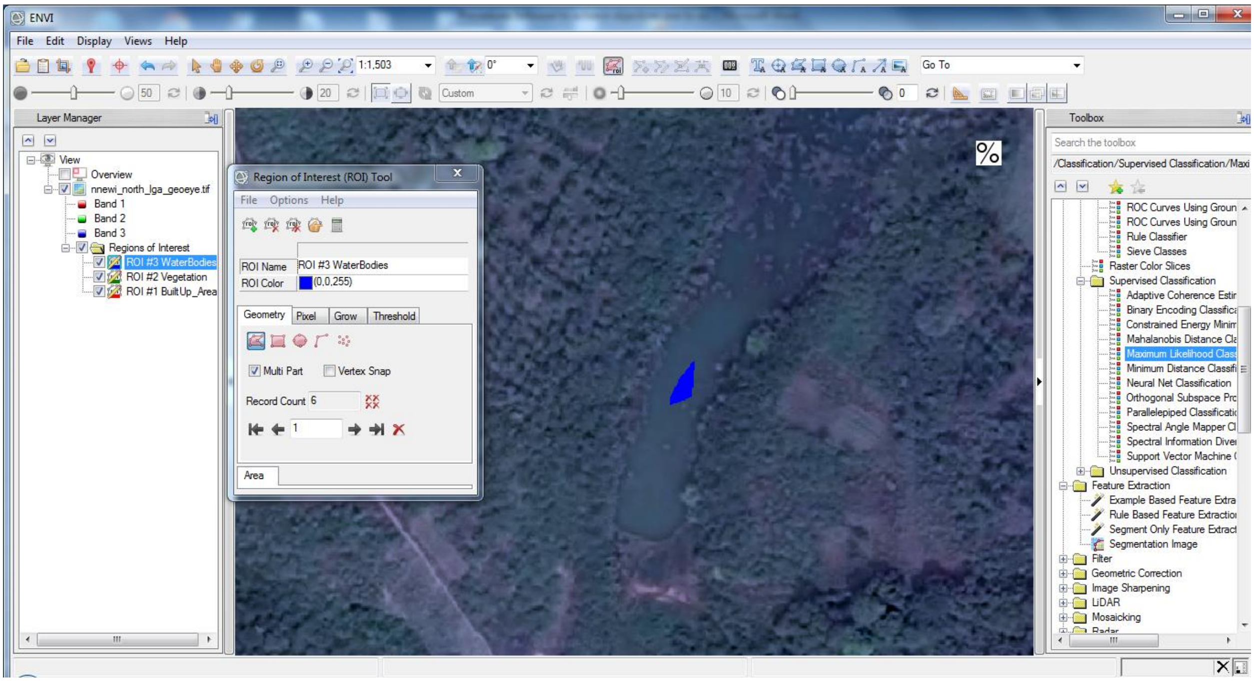


Figure 4.10: Water bodies FOI of trainingat Umudim

The ROI of collection of Water bodies at Umudim, this is a drainage basin that collect water annual, Fig.4.10.

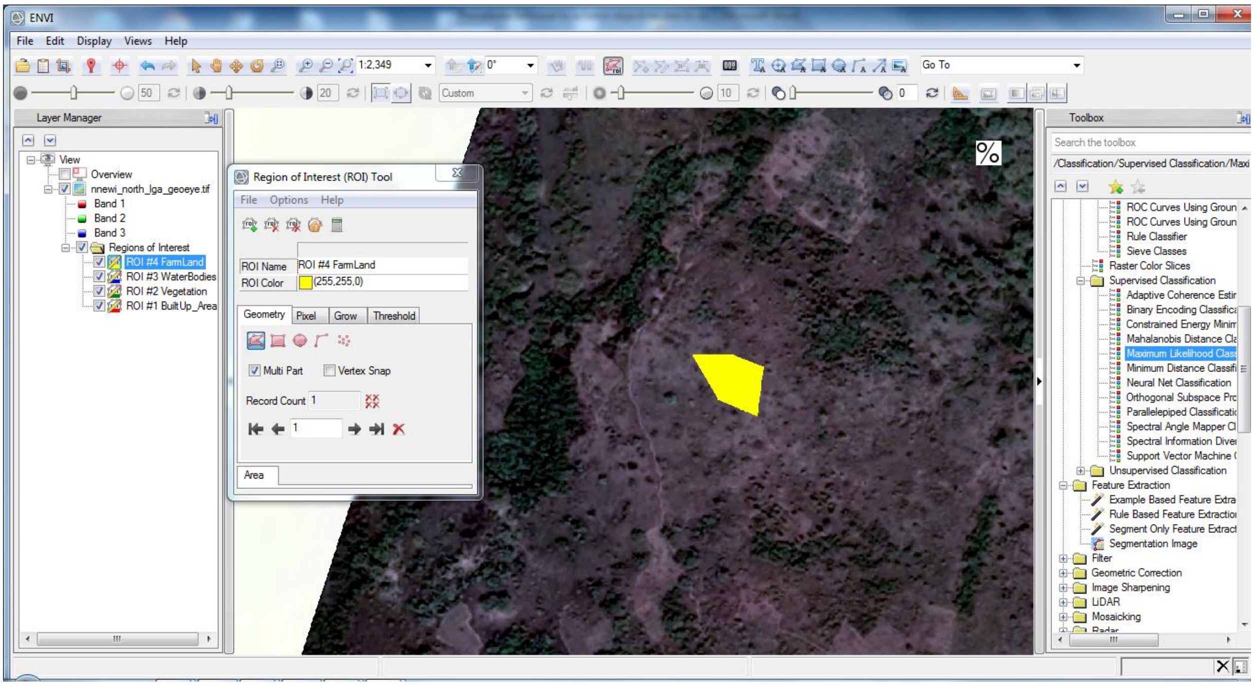


Figure 4.11: Farm Land as a FOI located towards the Ichi border of the City

FOIs Farm Land towards the outskirt of the City, Fig. 4.11, and suchlarge farm lands are noticeable in all the four quarters of Nnewi with little or no settlement around.

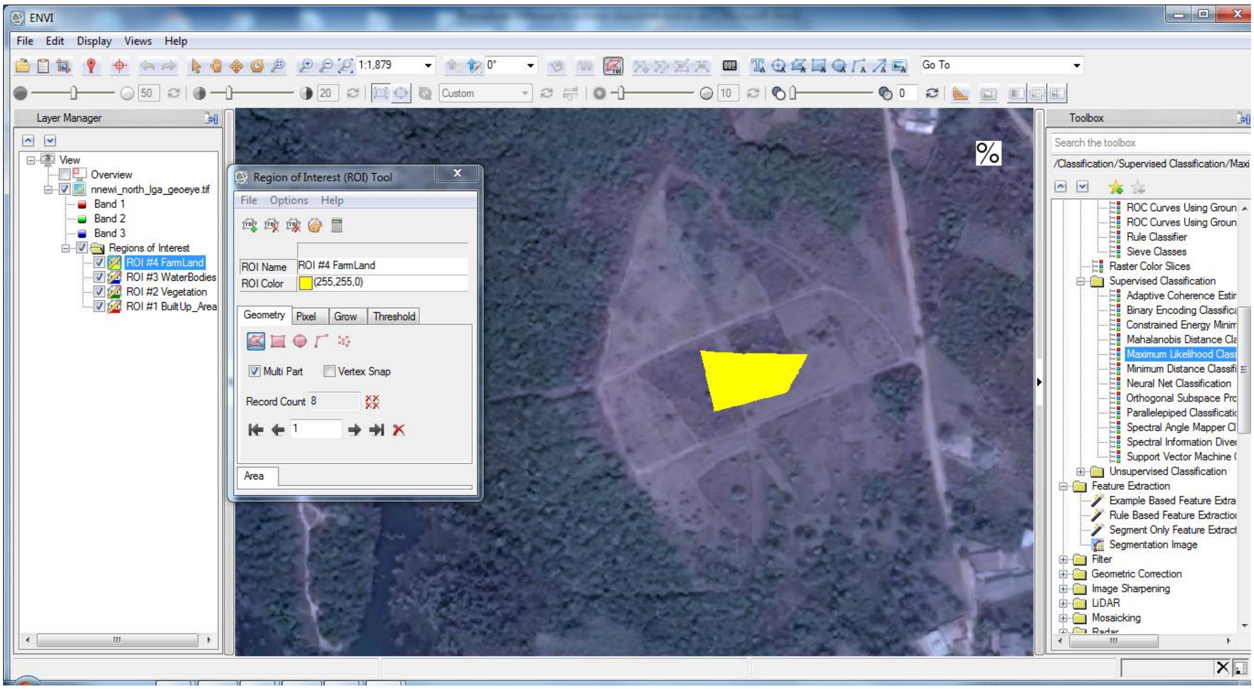


Figure 4.12: FOI of Farm Land with few settlement encroaching the area

ROIs Farm Land visualized with few encroaching settlements, Fig. 4.12, linked with a footpath or unpaved narrow roads. The encroaching settlements are gradually taking the available farm land and vegetation which is clear sign of urban-growth and population increase.

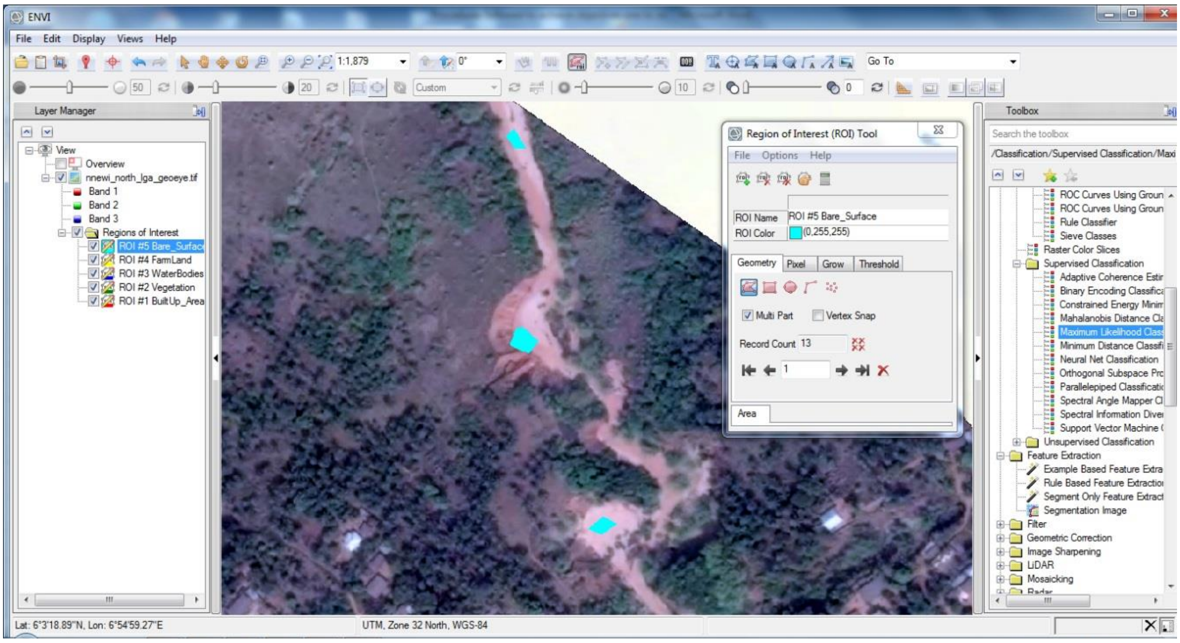


Figure 4.13: FOI of Open/bare Surface Land,exampleof erosion site at Nnewichi-Nnewi

FOI of Open/bare Surface area showing erosion site at Nnewichi-Nnewi, Fig. 4.13. This site is enormous and hazardous which had in 2017 attracted the attention of Nigerian Erosion and Watershed Management Project (NEWMAP) and in collaboration of World Bank and Anambra State Government.

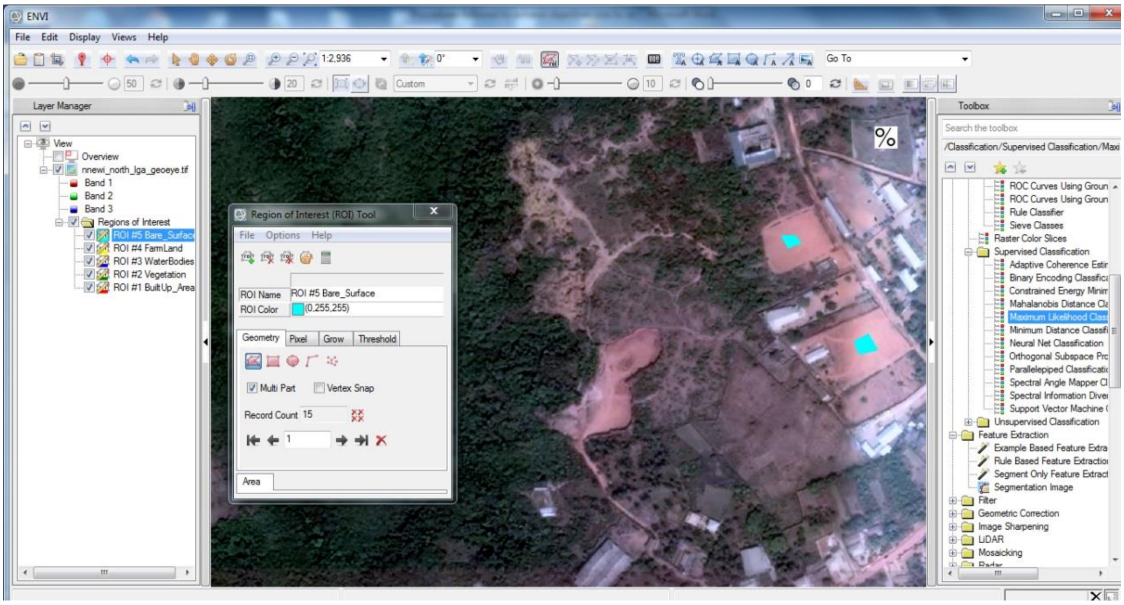


Figure 4.14: Example of Open/bare Surface LandFOI

This ROI is showing Open/bare Surface area, this is a typical example of open field showing playgrounds and sand mining site with zero vegetation,Fig.4.14.

4.2. Procedures Adopted in Analysing Objective Number Two.

Objective No. (2).“To evaluate the result of region of interests and ground truth extraction using the Jeffries-Matusita and Transformed Divergence separability index”was analysed as follows:

4.2.1. ROI separability index analysis

The research used the n-D Visualizer and ROI separability tool to compute and check the spectral separability between selected ROI pairs for a given input file. Both the Jeffries-Matusita and Transformed Divergence separability measures are testified. These values range from 0 to 2.0 and indicate how well the selected ROI pairs are statistically separated. Values greater than 1.9 indicate that the ROI pairs have good separability. For ROI pairs with lower separability values, you should attempt to improve the separability by editing the ROIs or by selecting new ROIs. For ROI pairs with very low separability values (less than 1), you might want to combine them into a single ROI.

The results of the whole statistical analysis are shown in Fig. 4.15 to 4.25; Both the Jeffries-Matusita and Transformed Divergence separability analysis were used in achieving the stated objective.

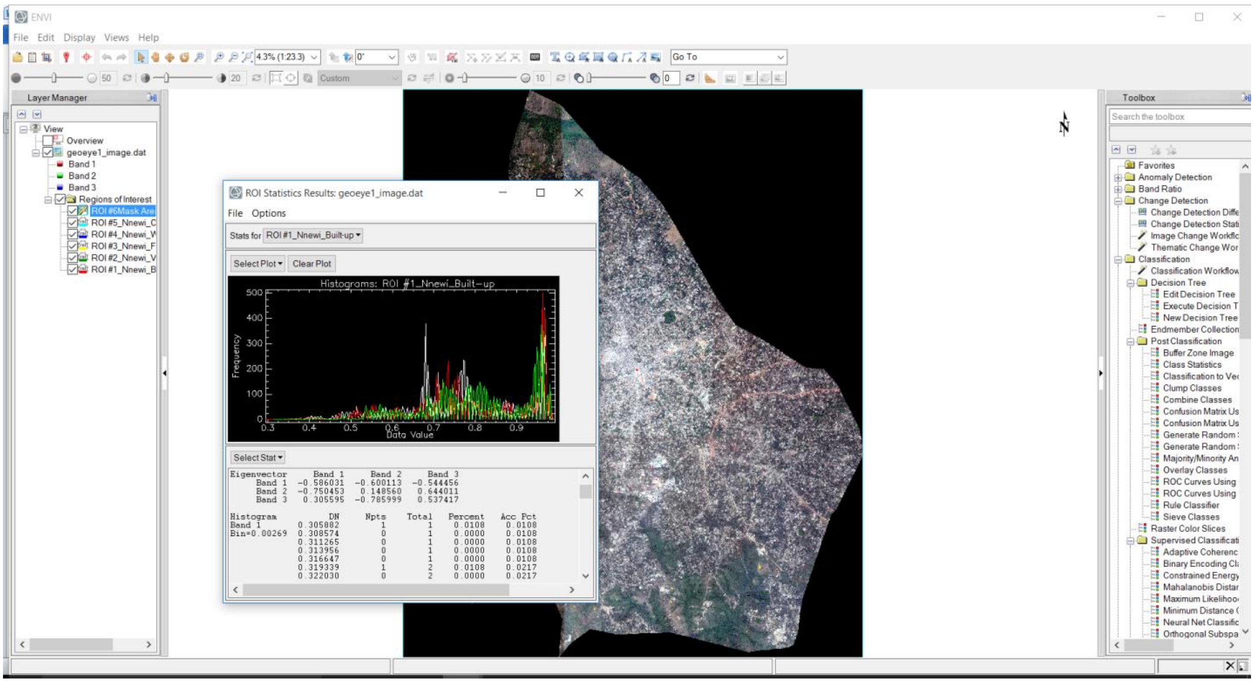


Figure 4.15: Features of interest statistic of built up area, the histogram of the three bands (RGB) used

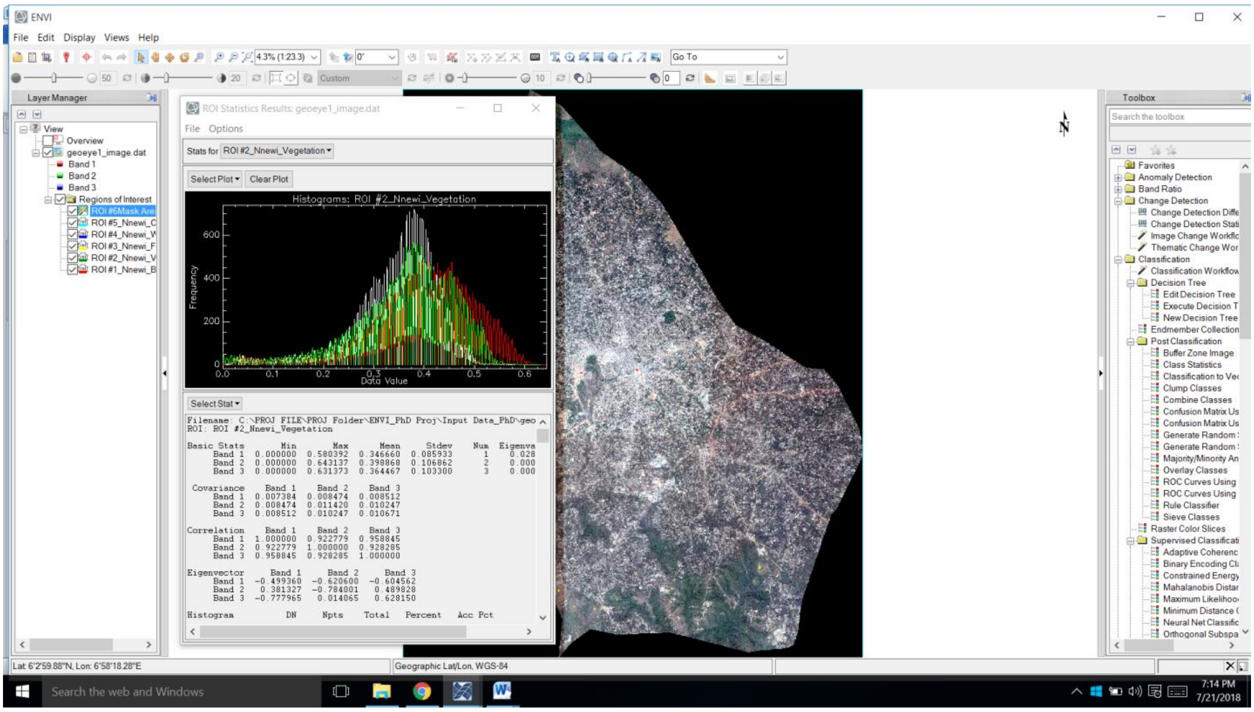


Figure 4.16: Features of interest statistic of vegetation area, the histogram of the three bands (RGB) used.

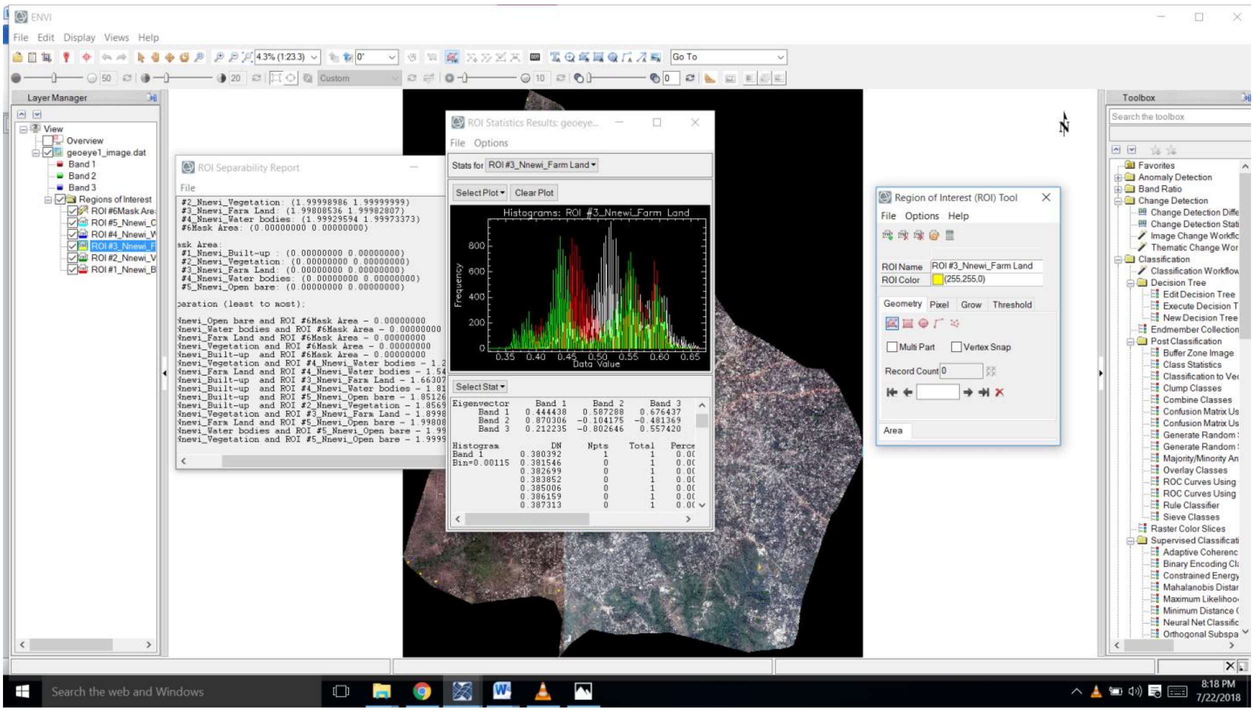


Figure 4.17: Features of interest statistic of farm land,the histogram of the three bands (RGB) used.

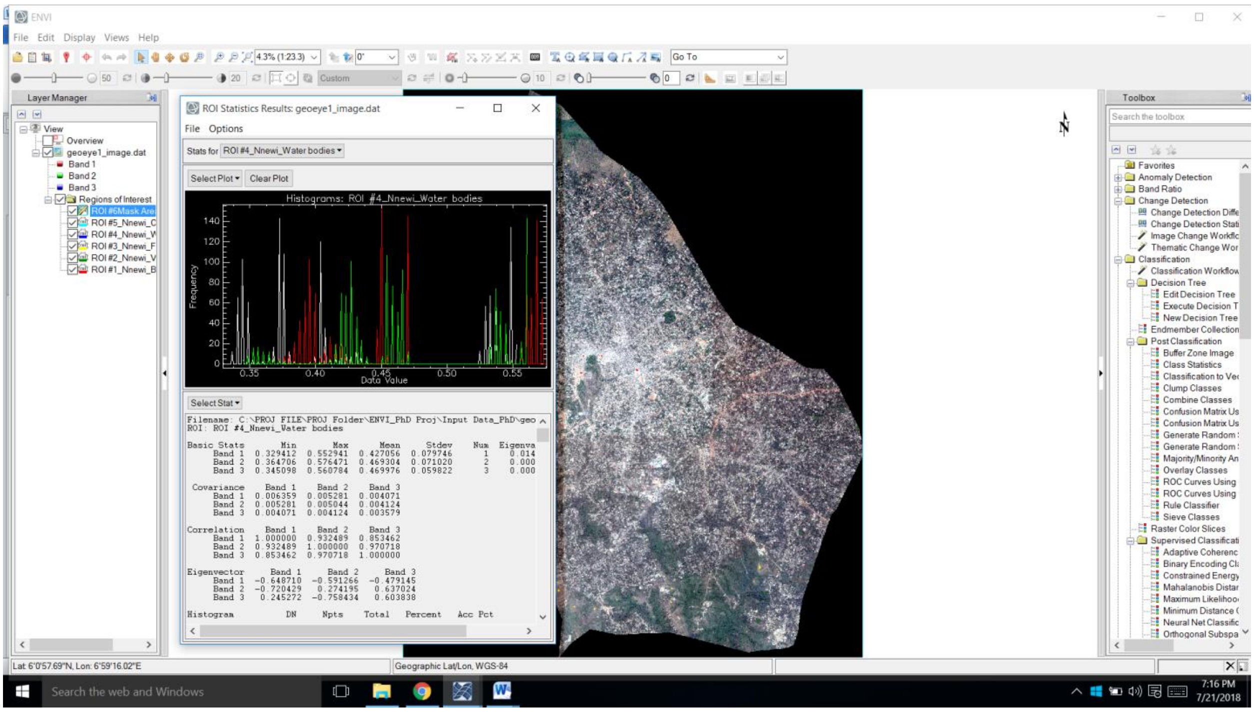


Figure 4.18: Features of interest statistic of water bodies, the histogram of the three bands (RGB) used

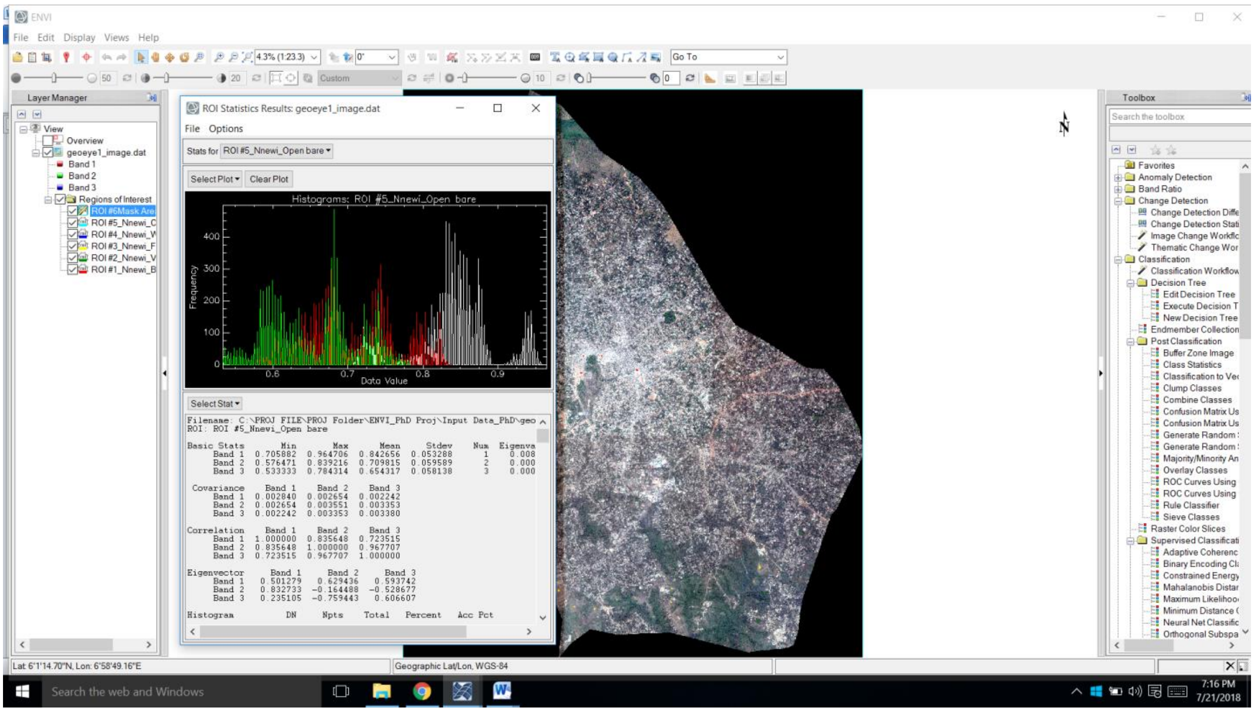


Figure 4.19: Features of interest statistic of open bare surface, the histogram of the three bands (RGB) used

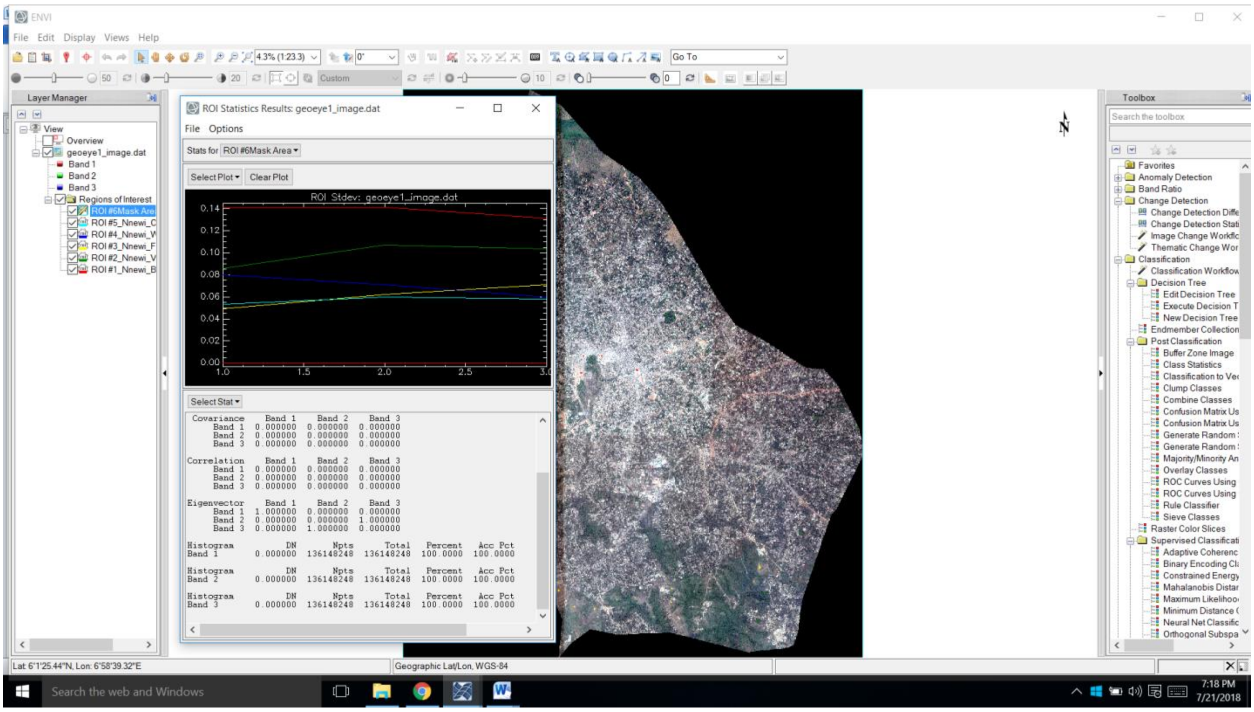


Figure 4.20: Result of features of interest statistic, the “Standard Deviation” performance of all thefeatures of interestused

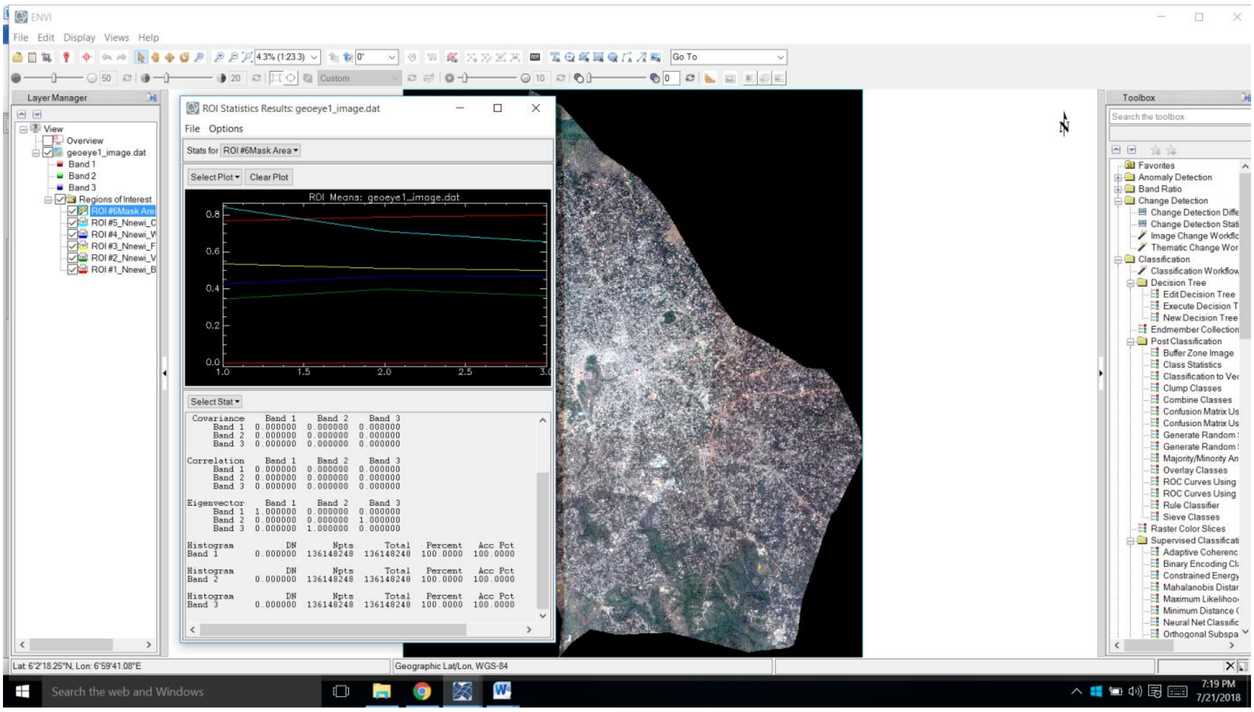


Figure 4.21: Result of features of interest statistic, the “Mean” performance of all the features of interestused

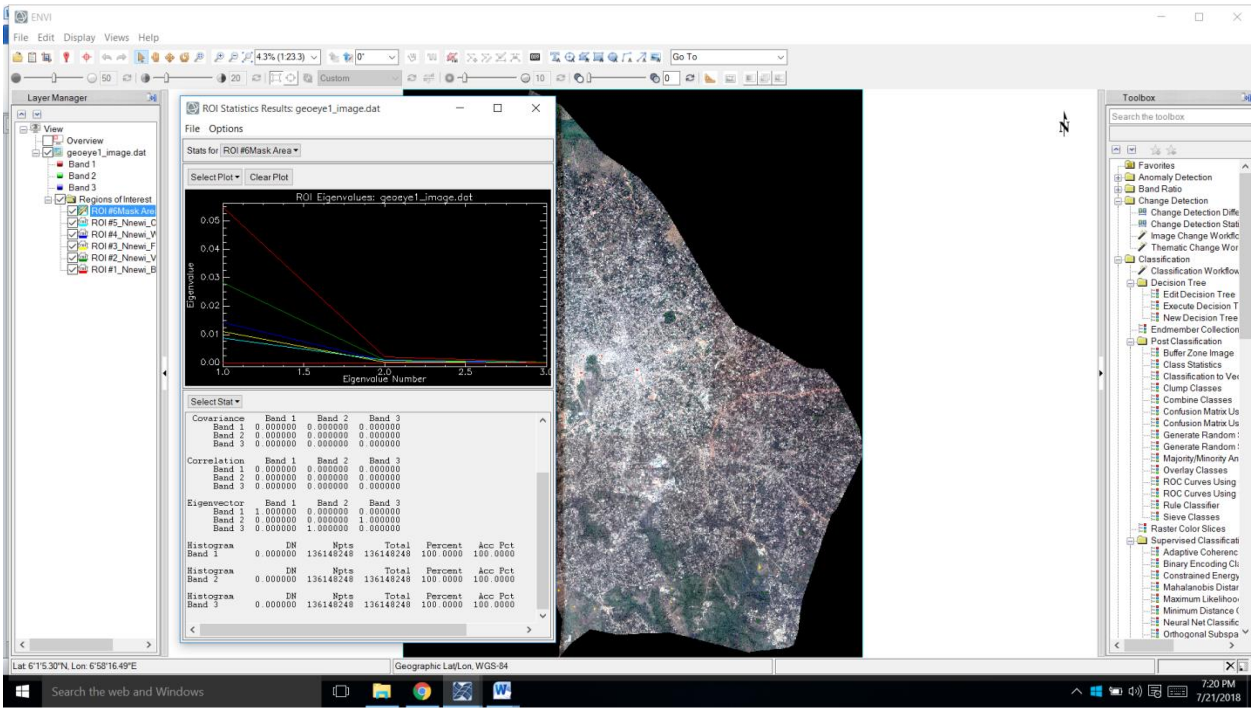


Figure 4.22: Result of features of interest statistic, the “Eigenvector” performance of all the features of interest used

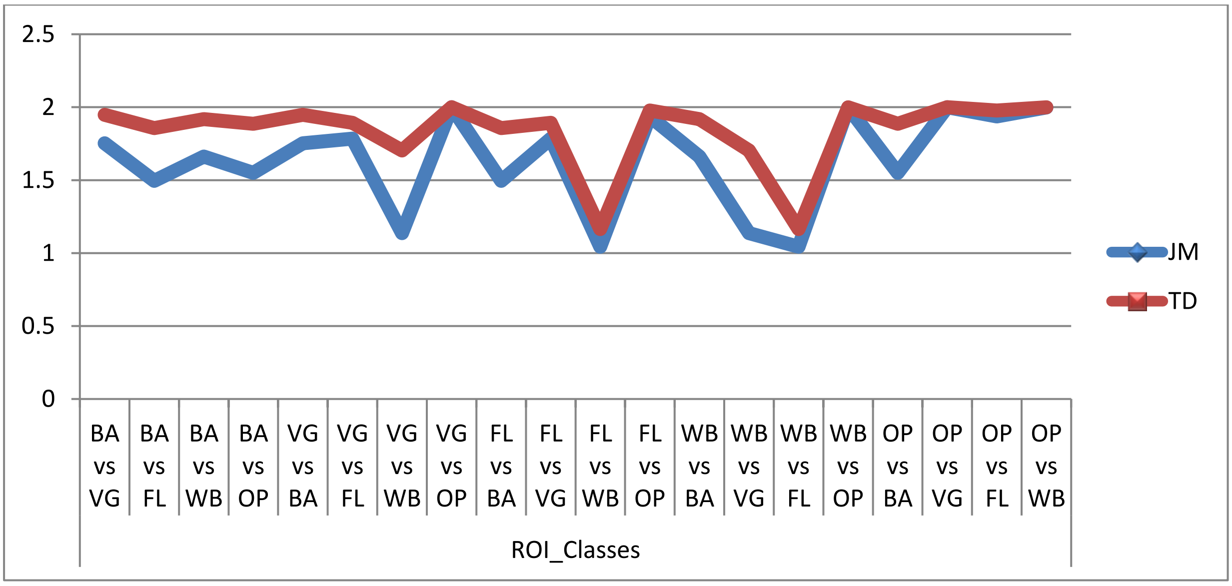


Figure 4.23: Jeffries-Matusita and Transformed Divergence separability of the Ground Truth ROIs. Where, JM means Jeffries-Matusita, while TD is Transformed Divergence. Similarly, BA is built-up area, VG isvegetation, FL is farm land, WB is water bodies, and OP is open/bare surfaces. The masked area was not used in computation because the value is zero.

The analysis was done such that one ROI (Class) was run against another and vice-versa to ascertain their separability.

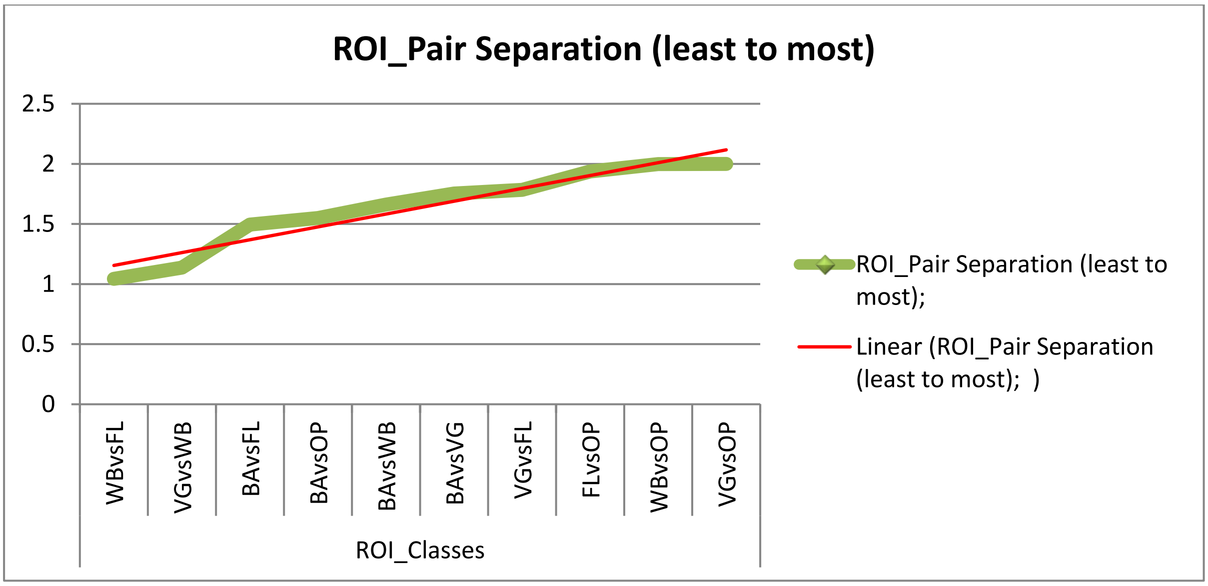


Figure 4.24: Graph of the Ground Truth ROIs pair separation (Least to most).

Where, BA is built-up area, VG isvegetation, FL isfarm land, WB iswater bodies, OP is open/bare surfaces. The masked area was not used in computation because the value is zero. The analysis was done such that one ROI (Class) was run against another and vice-versa to ascertain their separability.

Figure 4.25: Jeffries-Matusita and Transformed Divergence separability of the "Trained ROIs".

Where, JM means Jeffries-Matusita, while TD is Transformed Divergence. Similarly, BA is Built-up Area, VG is Vegetation, FL is Farm Land, WB is Water bodies, OP is Open/bare surfaces and While the masked area was not used in computation because the value is zero. The analysis was done such that one ROI (Class) was run against another and vice-versa to ascertain their separability, Figure 4.25.

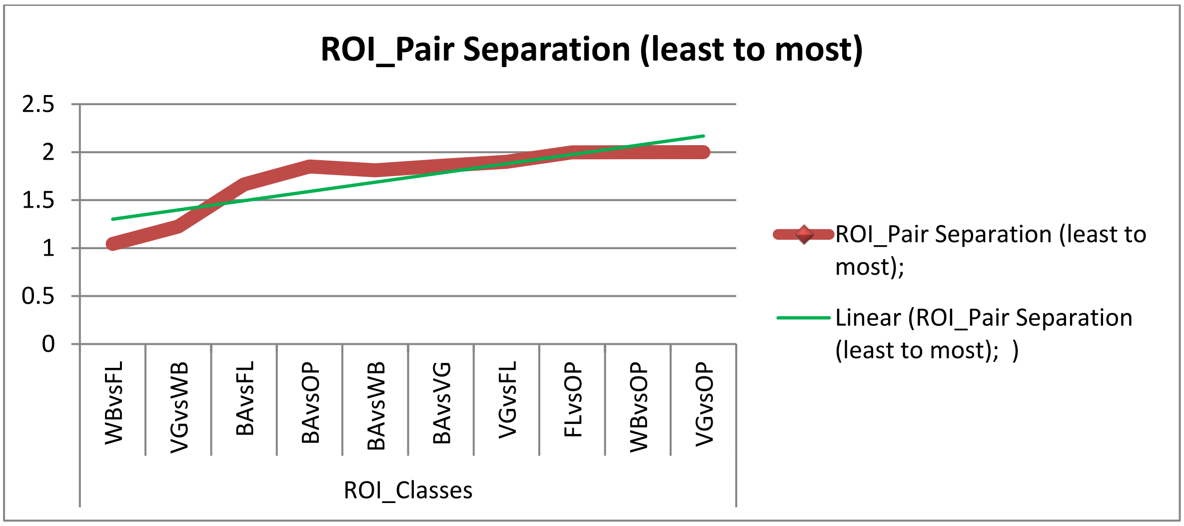


Figure 4.26: Graph of the "Trained ROIs" pair separation (Least to most)

Where, BA isbuilt-up area, VG is vegetation, FL isfarm land, WB is water bodies, OP is open/bare surfaces where the masked area was not used in computation because the value is zero. The analysis was done such that one ROI (Class) was run against another and vice-versa to ascertain their separability.

The whole statistical analysis in (Fig. 4.15 to 4.22) was targeted at achieving ROIs separability analysis shown in Fig. 4.23 to 4.26, where the Jeffries-Matusita and Transformed Divergence separability measures are tested as stated. The results of "Trained" and "Ground Truth" ROI indicated how well the selected ROI pairs are statistically separated. The values range from above 1 to greater than 1.9 which indicates that the ROI pairs have good separability. However, the summary of the separability index analysis shows that the Transformed Divergence (TD)performs better than the Jeffries-Matusita (JM) using either the trained ROIs or Ground truth ROIs.Objective of this section of research was achieved and results demonstrated the best way of going about it and why it is important to validate the areas of classes the analyst trained before using them for supervised image classification.

4.3. Procedures Adopted in Analysing Objective Number Three.

Objective No. (3) “To perform supervised classification using SVM and ML in ENVI Software” was analyzed as follows:

4.3.1.Launch the software and input the raw satellite image in the system

Once the downloaded raw satellite image is saved into a computer and the image is linked to a project folder, you can launch the ENVI software by double clicking on ENVI+IDL icon. it will automatically start the ENVI+IDL which is asoftware and programming language that contains an extensive library of image processing and analysis routines. With IDL, you can quickly visualized image data and begin investigating the best way to extract useful information.

4.3.2. Pre-processing of the image

The initial preparation for the processing of any remotely sensed image data is done under pre-processing. This involves reading the user requirements and the input data and storing these information & data for further processing in the appropriate format under a project area. The input data was radiometrically and geometrically corrected by GeoEye image cooperation. Data georeferencing and transformation was also carried by vendors. The study used ENVI (Environment for visualizing images) software version 5.1 for image processing, image enhancement, Filtering and masking.

4.3.3.Design image featureof interests(FOIs)

FOIs are usually image classification scheme containing information of classes of interest such as built up area, farm lands, vegetation areas, water bodies, and open/bare surface among other features of interest. Supervised classification requires that the analyst be familiar with the area of interest and needs to know where to find these classes of interest in the area covered by the satellite image. The “Extraction ROIs” was careful carried out by creating ROIs from HRSI geometry usingROIs tool. Region of interest (ROI) uses a point, polyline, or polygon object drawn on an image, used it to define a specific area of interest for extracting classification statistics, masking, and other operations in ENVI are possible. From a processing standpoint, ROIs are pixel addresses with associated data. With the assistance of the ROI editor/ toolbar, you can easily create training samples to represent the classes you want to extract. For the computer to be properly guided on the classification process the selected ROIs are assigned with name, codes and colour. The five classes of interest earlier adopted which includes the built-up areas; vegetation, wetlands/water bodies, buildings/pavement, open/bare surfaces and farm lands were extracted and saved in different format. More Details in subsection 4.2.

4.3.4. Selecting representative areas of the image and generating training signatures.

The Trained ROIs/Ground Truth ROIs was created (Fig.4.27 and Fig.4.28) and used to produce ground truth image that used for confusion matrix (contingency matrix) analysis.

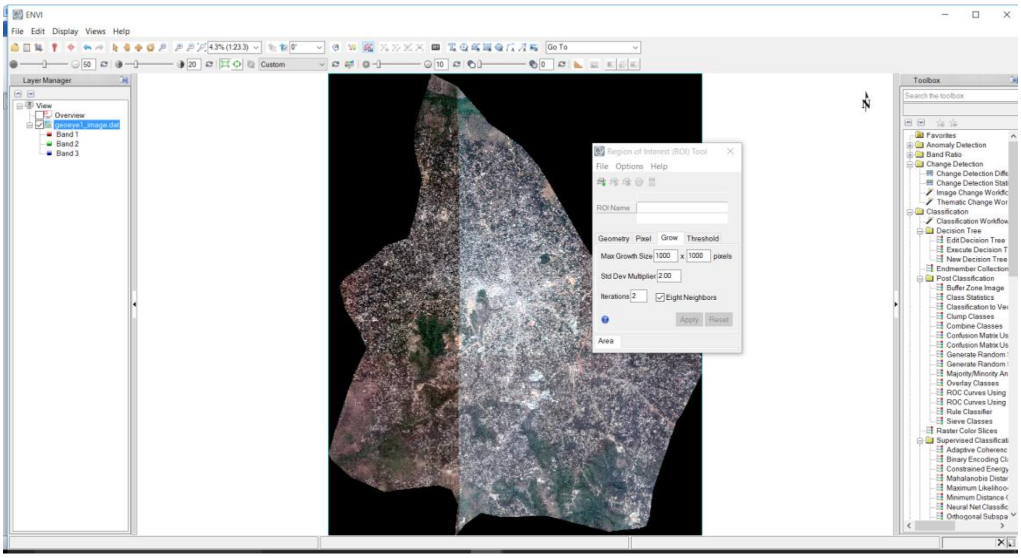


Figure 4.27: shows the ROI tool in action during extraction of ROI classes.

Fig 4.27 is showing ROI tool in action during extraction of ROI classes., the analyst used #1 to #5 as the code for the five classes of interest and various colours that can distinguish one class from another and the colour selected were close to the natural colour of the features of interest.

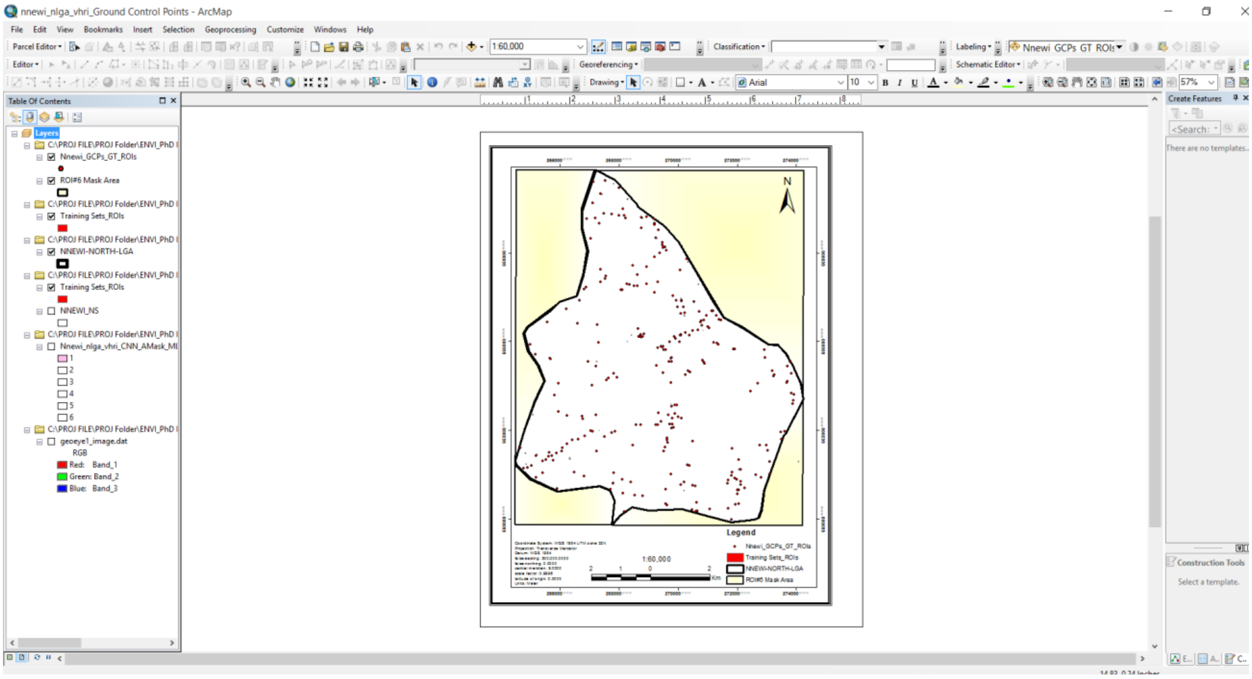


Figure 4.28: Shape file of the Trained FOIs/Ground Truth ROIs that was used

4.3.5. Run image classification algorithms

Image classification refers to the computer-assisted interpretation of remotely sensed images. ENVI software version 5.1 was used for GeoEye-1 image supervised classification. The main classifier used was SVM and its results were compared using ML.Object-based processing was adopted which are techniques that classify a set of input objects rather than classifying pixels individually. Object is a region of interest with spatial, spectral, and/or texture characteristics (brightness, colour, etc.) that define the region.

SVM and ML under the same condition of ENVI window was tested, see Fig.4.29. SVM at its 99% supervised classification stage andFig. 4.30; ML at its 98%supervised classification stage and consequently Figure 4.27 shows SVM at its 100% Classification Process.

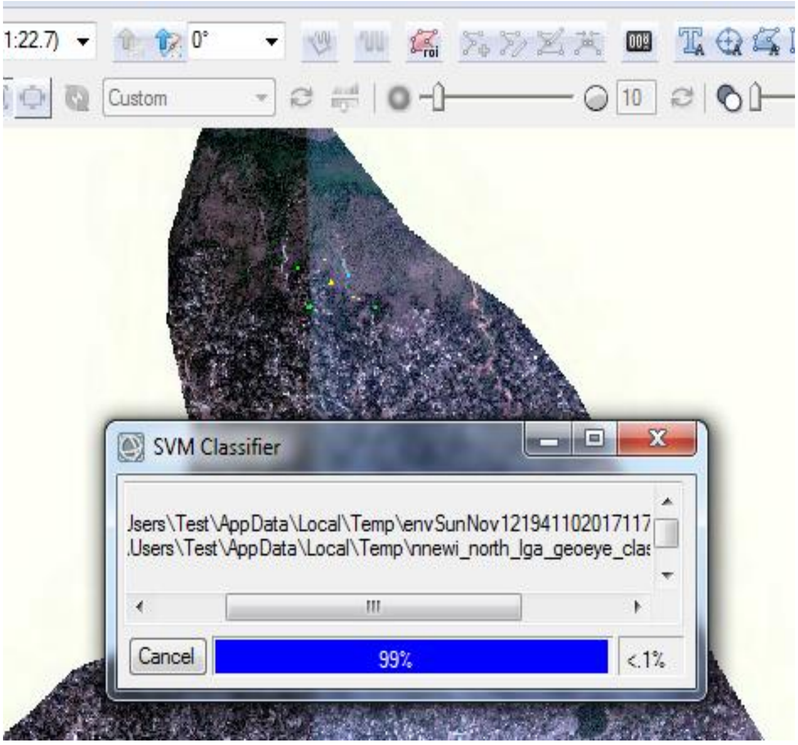


Figure 4.29: SVM at its 99% Supervised Classification Stage

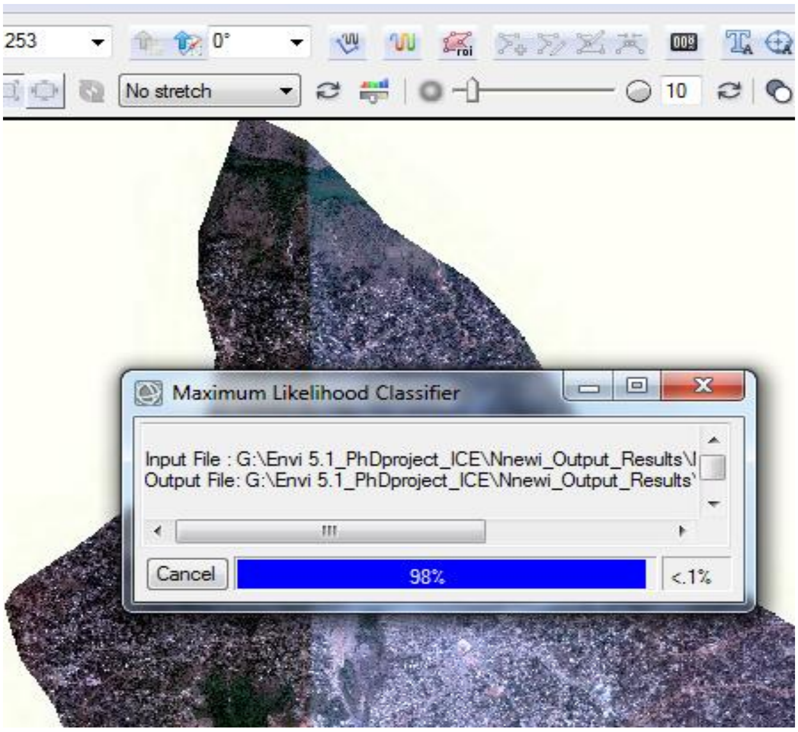


Figure 4.30: ML at its 98% Supervised Classification Stage

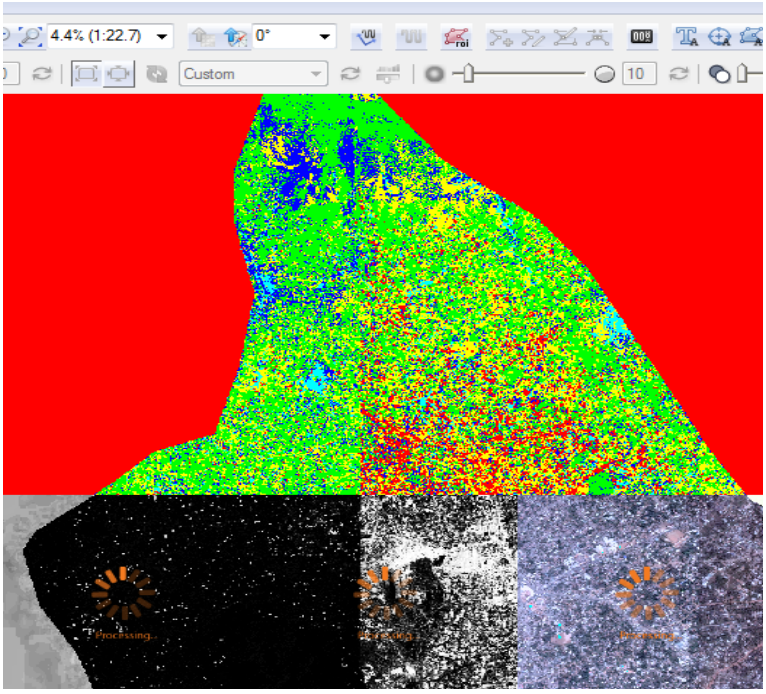


Figure 4.31: SVM at its 100% Supervised Classification Process

Fig.4.31. is showing the transformation of the satellite image from its raw RGB stage to classified colourful landuse/landcover raster image phase.

4.3.6.Post-processing

In ENVI,post classification tools were used to classify rule images, to calculate class statistics and confusion matrices, to apply majority or minority analysis to classification images, to clump, sieve, and combine classes, to overlay classes on an image, to calculate buffer zone images, to calculate segmentation images, and to output classes to vector layers. In Post-processing analysis, ENVI was used to calculate a confusion matrix (contingency matrix) using ground truth ROIs. In each case, an overall accuracy, producer and user accuracies, with their errors of commission and omissionand kappa coefficient were computedand presented using aconfusion matrix table.

The studywas not satisfactorywith the performance of SVM and ML in mapping geometric satellite features using only overall accuracies and Kappa Statistical Procedures and therefore developed more stimulating testing parameters from the outcome of the contingency matrix analysis, kappa coefficient and the behavours of these classifiers during image classification. The developed stimulatingformula is called “Post Confusion Matrix” (PoCoMa). This formula was formulated from the contingency matrix, kappa coefficient and the behaviours of the image classifiers and tested for computing and comparing the capabilities of image classifiers. The values obtained were tested using t-statistics (Paired Samples Test).

4.3.6.1. *Calculation of Confusion Matrices*

Confusion matrix (Also called a contingency matrix.) is a statistical table used to assess classification accuracy and misclassification between categories. The matrix is size m x m, where m is the number of classes. The rows in the matrix represent classes that are assumed to be true, while the columns represent classes derived from remote sensing imagery. The matrix also lists errors of commission and omission. Confusion matrix can be used to show the accuracy of a classification result by comparing a classification result with ground truth information. ENVI can calculate a confusion matrix (contingency matrix) using either a ground truth image or using ground truth ROIs. In each case, an overall accuracy, producer and user accuracies, kappa coefficient, confusion matrix, and errors of commission and omission are reported (the result of confusion matrix used in this section was computed during classification validation, see table 4.1 and 4.4.

4.3.7*Accuracy assessment compares classification results with field studies.*

The final classes must be assessed to ensure it is within permissible accuracy. This is done by generating confusion matrix and calculating derived errors and their accuracies. Also several sites were visited (ground truthing) in the study area with respect to the adopted classes of land use and land cover. Coordinates of the sites were also taken. Reference datacollected through the process of “ground truthing”, determination of class types at specific locations and Compare reference to classified map. Having done that you have to ask yourself, does class type on classified map = class type determined from reference data? If yes, it is ok and if not, then the analyst has to go back to drawing board because the goals of accuracy assessment are to assess how well a classification worked and to understand how to interpret the usefulness of someone else's classification, (Table 4.8 and Equation. 4.1).

4.3.7.1. *Overall accuracy*

This is computed by dividing the total correct number of pixels (i.e. summation of the diagonal) to the total number of pixels in the matrix (grand total). The overall accuracies for the SVM map is (328811760/335285600) = 98.0692% and while that of ML is = (276605139/335285600) = 82.4984%.

4.3.7.2. *Producer's accuracy*

This refers to the probability of a reference pixel being classified correctly. It is also known as omission error because it only gives the proportion of the correctly classified pixels. It is obtained by dividing the number of correctly classified pixels in the category by the total number of pixels of the category in the reference data. The overall result of the producer's accuracy ranges from 91.10% to 100% for SVM while ML is 45.04% to 85.13%. The result of SVM was adopted but ML was rejected because the lowest producer's accuracy exists in the land cover classes of "water bodies" and "open/bare surface". This is probably attributed to the similar spectral properties of some of the land cover classes (for instance, vegetation and farm land with urban areas and vice versa or the noise was not removed using MLC.

4.3.7.3. *User's accuracy*

This assesses the probability that the pixels in the classified map or image represent that class on the ground (Congalton, 1991). It is obtained by dividing the total number of correctly classified pixels in the category by the total number of pixels on the classified image. User's accuracy of individual classes ranges from 89.76% to 100% for SVM while ML is 45.58% to

96.97%. The result of SVM was adopted but ML was rejected because from ML user's accuracy point of view, vegetation andwater bodies presented low accuracy for the land cover map. The vegetation and water bodies were, to some extent, misclassified as Built-up areas, vegetation and farm lands respectively. This is probably caused by the inconsistent spectral signature of the featuresor the noise was not removed using ML.

4.3.7.4. *Kappa coefficient*

The kappa coefficient determines the difference between the observed agreement between two maps and the agreement that might be attained by change matching of the two maps. It expresses the proportionate reduction in error generated by a classification process compared with the error of a completely random classification (Congalton, 1991). The Kappa statistic incorporates the off-diagonal elements of the error matrices (i.e., classification errors) and represents agreement obtained after removing the proportion of agreement that could be expected to occur by chance. Error Matrix and Kappa Coefficient results revealed that SVM is better than ML as follows (SVM overall Accuracy is 98.07% and Kappa Coefficient is 0.97 while ML overall Accuracy is 82.50% and Kappa Coefficient is 0.76. These Kappa results are considered to be a good result because an interpretation for SVM is between (0.81-1.00) which is almost perfect agreement, whileKappa Coefficient for ML is between (0.61-0.80) which is a substantial agreement.

Table 4.1: Confusion Matrix (Contingency Matrix) profileof SVM

SVM Classified Classes	**Ground Truth**									
	Unclassified	**Masked Pixels**	**Built-up**	**Vegetation**	**Water bodies**	**Farmland**	**Open space**	**Total**	**Error of Commission**	**User's accuracy**
Unclassified	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0.00**
Masked Pixels	**0**	**137951157**	**0**	**0**	**0**	**0**	**0**	**137951157**	**0**	**100.00**
Built-up	**0**	**0**	**24020339**	**0**	**39843**	**1576383**	**1051279**	**26687844**	**0.17**	**99.83**
Vegetation	**0**	**0**	**1015532**	**21779789**	**23473652**	**1519190**	**0**	**47788163**	**10.24**	**89.76**
Water bodies	**0**	**0**	**1325752**	**3895335**	**30825253**	**7467781**	**0**	**77742135**	**1.50**	**98.50**
Farmland	**0**	**0**	**0**	**371737**	**0**	**43142384**	**0**	**43514121**	**0.85**	**99.15**
Open space	**0**	**0**	**0**	**0**	**0**	**0**	**1602180**	**1602180**	**0.00**	**100.00**
Total	**0**	**137951157**	**27047015**	**44475019**	**76854936**	**47355293**	**1602180**	**335285600**		
Error of Omission	**0**	**0**	**1.50**	**3.55**	**0.36**	**8.90**	**0.00**			
Producer's accuracy	**0.00**	**100.00**	**98.50**	**96.45**	**99.64**	**91.10**	**100.00**			

Overall Accuracy (328811760/335285600) =98.07%, Kappa Coefficient = 0.97.

Table 4.2: Support Vector Machineprofile of masked area of the study area

S/N	**Classes (ROIs)**	**Code**	**Colour Used**	**Points (PIXELS)**	**Percentage %**	**Area (km²)**
1	Unclassified	0	Black	0.000	0.00000	0.00
2	Built-Up Area	#1	Mars Red	26,687,844	7.960	929.00
3	Vegetation	#2	Leaf Green	47,788,163	14.253	1,663.51
4	Water Bodies	#3	Lapis Lazuli	43,514,121	12.978	1,514.73
5	Farm Land	#4	Solar Yellow	77,742,135	23.187	2,706.20
6	Open/Bare	#5	Indicolite Green	1,602,180	0.478	55.77
7	Masked Pixels	#6	White/Pink	137,934,677	41.139	4,801.51
Total				335,285,600	99.995	11,670.72

Table 4.3: Support Vector Machineprofile of unmasked area of the study area

S/N	Classes (ROIs)	Code	Colour Used	Points (PIXELS)	Percentage %	Area (km²)
1	Built-Up Area	#1	Mars Red	26,687,844	13.52	929.00
2	Vegetation	#2	Leaf Green	47,788,163	24.23	1,663.51
3	Water Bodies	#3	Lapis Lazuli	43,514,121	22.05	1,514.73
4	Farm Land	#4	Solar Yellow	77,742,135	39.40	2,706.20
5	Open/Bare	#5	Indicolite Green	1,602,180	0.80	55.77
		Total		197,334,443	100.00	6,869.21

Table 4.4: Confusion Matrix (Contingency Matrix) profile of ML

MLC Classified Classes	Ground Truth (Pixels)								
	Masked Pixels	Built-up	Vegetation	Water bodies	Farmland	Open space	Total	Error of Commission	User's accuracy
Masked Pixels	137951157	0	0	0	0	0	137951157	0	100.00
Built-up	0	24020339	0	39843	1576383	1051279	26687844	10	90
Vegetation	0	1015532	21779789	23473652	1519190	0	47788163	54.42	45.58
Water bodies	0	1325752	3895335	30825253	7467781	0	43514121	29.16	70.84
Farmland	0	3169737	0	14096337	60474995	1066	77742135	22.21	77.79
Open space	0	48574	0	0	0	1553606	1602180	3.03	96.97
Total	137951157	29579934	25675124	68435085	71038349	2605951	335285600		
Error of Omission	0	18.8	15.17	54.96	14.87	40.38			
Producer's accuracy	100.00	81.20	84.83	45.04	85.13	59.62			

ML: Overall Accuracy = (276605139/335285600) 82.4984% and Kappa Coefficient = 0.7626

Table 4.5:MLresults includingthe masked area of the study area

S/N	Classes (ROIs)	Code	Colour Used	Points (PIXELS)	Percentage %	Area (km²)
1	Unclassified	0	Black	0.000	0.00000	0.00
2	Built-Up Area	#1	Mars Red	29,579,934	8.822	1029.678
3	Vegetation	#2	Leaf Green	25,675,124	14.253	893.751
4	Water Bodies	#3	Lapis Lazuli	68,435,085	12.978	2382.225
5	Farm Land	#4	Solar Yellow	71,038,349	21.187	2472.845
6	Open/Bare	#5	Indicolite Green	2,605,951	0.777	90.713
7	Masked Pixels	#6	White/Pink	137,951,157	41.144	4802.079
		Total		335,285,600	99.999	11,671.291

Table 4.6: ML, image classification profile of unmasked area of the study area

S/N	Classes (ROIs)	Code	Colour Used	Points (PIXELS)	Percentage %	Area (km²)
1	Built-Up Area	#1	Mars Red	29,579,934	14.98	1,029.678
2	Vegetation	#2	Leaf Green	25,675,124	13.01	893.751
3	Water Bodies	#3	Lapis Lazuli	68,435,085	34.68	2,382.225
4	Farm Land	#4	Solar Yellow	71,038,349	36.00	2,472.845
5	Open/Bare	#5	Indicolite Green	2,605,951	1.32	90.713
		Total		197,334,443	100.00	6,869.21

Overall Accuracy = (207239/240806) =82.50%, Kappa Coefficient = 0.76.

4.3.8. Production of maps and results from the classification image

ArcGIS 10.1 was employed for producing thematic maps. Microsoft Excel, GraphPad Prism and SPSS were employed for data analysis and results presentation.

One of the justifications of this research is lack of adequate and correct geospatial information of Nnewi-North LGA to aid daily planning, development, monitoring and accurate decision-making. Example is illustrated in (Appendix A1)which is showing old Nnewi Local Government Area map comprising of some part of Ekwusigo LGA that are most commonly used for today administrative planning, development, monitoring and decision-making.Since the history of States and Local government area creations in Nigeria in 1967, 1976,1987, 1991and 1996 respectively some of our administrative boundary are not updated and one of them is the map of old Nnewi Local Government Area map(Appendix A1). Thus, this research work sought and used subset of the satellite imagery covering Nnewi-North and South LGAin producing a baseline landuse and landcover map of study area using the correct shape file of Nnewi-North, (Appendix A2). To represent the correct land extent of Nnewi-North LGA after Ekwusigo LGA was carved out of Nnewi-North in 1996 by General Sani Abacha (GCFR).

Hence the research objective No. 5 “To produce the landcover map of Nnewi-North L.G.A”, that will serves as a base map for improved land-use planning and monitoring by end-users was achieved with an additional advantage of using the most current boundary line of Nnewi-North L.G.A (AppendixA2),where the results reveals enormous evidence aboutthe present percentages and size of the built-up area, Vegetation cover, Water bodies, Farm land and Open/bare surface of the classes of interest classified with respect to thiscurrent boundary line.The results of the classified GeoEye-1 2016 Satellite image of Nnewi North using ENVI SVM and ML algorithms was modified and produced in ArcMap 10.2”. The land cover maps produced and other results are shown in figures 4.28 and 4.29, table4.7 and appendix A2, B1 to B4).

4.3.9. Result of supervised classification using SVM and ML in ENVI Software

The study also realised that Remote Sensing and statistical integration can provide valuable ways of regular exploring the process of Urban Development (UD) as well as offering a means of evaluating the environmental and social consequences of various planning scenarios towards sustainable development strategy and that ENVI Software is robust tool for automatic image identification and classification but requires subjective manipulation of input parameters.

What is the settlement pattern of Nnewi-North LGA? Classified result shows 60% of nucleated settlement pattern observed in the Central Business Distrect Area (CBDA) while the remanining 40% were dispersed settlement pattern scateredd across the Local Government area. The result indicates that land is becoming a scarce and valuable commodity, especially in the cities center, effective land use is therefore necessary for urban sustainability. Moreover, urban structures are dynamic because population and anthropogenic activities are in a constant process of change and growth and urbanization has an important influence on the spatial distribution of land use. Thus, below is the present situation of Nnewi landcover:

4.3.9.1. *Built-Up Areas*

Using SVM, Built-up Areas is 13.52%, while ML result shows that Built-up Areas is 14.99%, adopting the result of SVM, the percentage of built-up area shows that Nnewi-North LGA still has available land for good developmental planning, that is, since out of 100% its only 13.6% has been developed. Hence, urban renewal can recover some percentage out of 13.6% when the government is willing to plan for the future. It must start with urban renewal, proper zoning and coordination to have an adorable smart city that will attract the foreign investors to the city. Since, the statistical analysis reveals a clear indication of increase in population and infrastructure development in the urban area, there is no other option than to plan for the sustainable development (SD) of Nnewi.

4.3.9.2. *Vegetation*

The result of SVMis 24.23% of Vegetation, and ML shows that Vegetation is 13.01%. However, the result of SVM was adopted because theUser's accuracy is 89.76%which shows that the Errors of Omission and of Commission of SVM were less compared with the ML. The statistical results show that there are pressures on the land covers (the coefficient of the correlation of the hypothesis is 0.710 with a significance value of 0.007) and this means that there is a strong positive correlation between extent of population growth and extent of land ownership and use in Nnewi north LGA. This calls for proper planning, enforcement and periodic monitoring of the city. This can be as a result of upsurge demand for land for Residential, commercial services, industrial complexes, transportation among other infrastructures.

4.3.9.3. *Water Bodies*

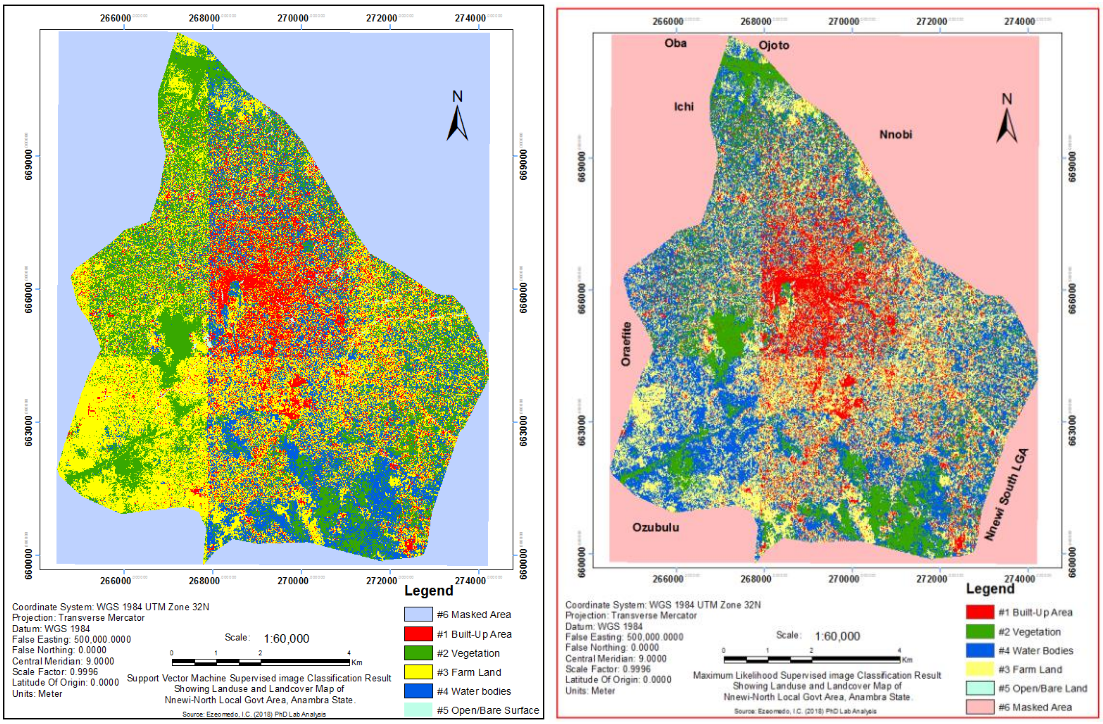
These include rivers, streams, lakes, etc. The 22.05% of Water bodies was computed for SVM and while using ML Water bodies = 34.08%. Conversely, the result of SVM was adopted because theUser's accuracy is 98.50%which shows that the Errors of Omission and of Commission of SVM were less compared with the ML. Artificial lakes were visualised around Umudim and Otolo, while other water bodies include Idemili River around Nnewi-ichi border with Ojoto, Ele River (Mmiri-Ele) within the Uruagu/ Umudim flowing East-west of the City and Ubu River flowing South-Eastern part of the Umudim-Otolo Nnewi within the border area of Ubu-Osigbu, Ukpor.

4.3.9.4. *Farm Land*

Includes, gardens, field crops, Horticulture, orchards, improve pasture, ploughed fields and fallow land as captured and trained by the analyst.Farm landis equal to 39.40% using SVMand Farm land is 36.00%using both ML. The result of SVM was adopted because the User's accuracy is 99.15% which shows that the Errors of Omission and of Commission of SVM were less compared with the ML.The contribution of Statisticalanalysis shows that erosion, flooding, lack of surplus of agricultural lands and poor soils are some environmental challenges facing Nnewi-North. Farm lands are noticeable around urban fringes of city within all the four quarters of Nnewi with little or no settlement around them.

4.3.9.5. *Open/Barren Land*

This class includes excavation sites, barren lands, erosion sites, playground and unpaved roads. This class was accurately classified using both SVM and ML because proportion of theirErrors of Omission and that of Commission of SVM were less. The percentage of open/barren land using SVM0.81% and while open/bare surfaceusing MLis 1.32%, figures 4.32 and 4.33, table 4.7, 4.8 and appendix B3



(a) SVM (b) ML

Figure4.32. Result of Image Classification Using (a) SVM (b) ML

Table 4.7: Percentage of Nnewi-North Land Use and Land Cover using SVM

Resultobtained using Support Vector Machine(SVM)

Classes (ROIs)	**Code**	**Colour Used**	**Points (PIXELS)**	**Percentage %**	**Area (km²)**
Built-Up Area	#1	Mars Red	26,687,844	13.52	929
Vegetation	#2	Leaf Green	47,788,163	24.23	1,663.51
Water Bodies	#3	Lapis Lazuli	43,514,121	22.05	1,514.73
Farm Land	#4	Solar Yellow	77,742,135	39.4	2,706.20
Open/Bare	#5	Indicolite Green	1,602,180	0.8	55.77

Table 4.8: Percentage of Nnewi-North Land Use and Land Cover usingML

Result obtained using Maximum Likelihood (ML)

Classes (ROIs)	**Code**	**Colour Used**	**Points (PIXELS)**	**Percentage %**	**Area (km²)**
Built-Up Area	#1	Mars Red	29,579,934	14.99	1,029.68
Vegetation	#2	Leaf Green	25,675,124	13.01	893.751
Water Bodies	#3	Lapis Lazuli	68,435,085	34.68	2,382.23
Farm Land	#4	Solar Yellow	71,038,349	36	2,472.85
Open/Bare	#5	Indicolite Green	2,605,951	1.32	90.713

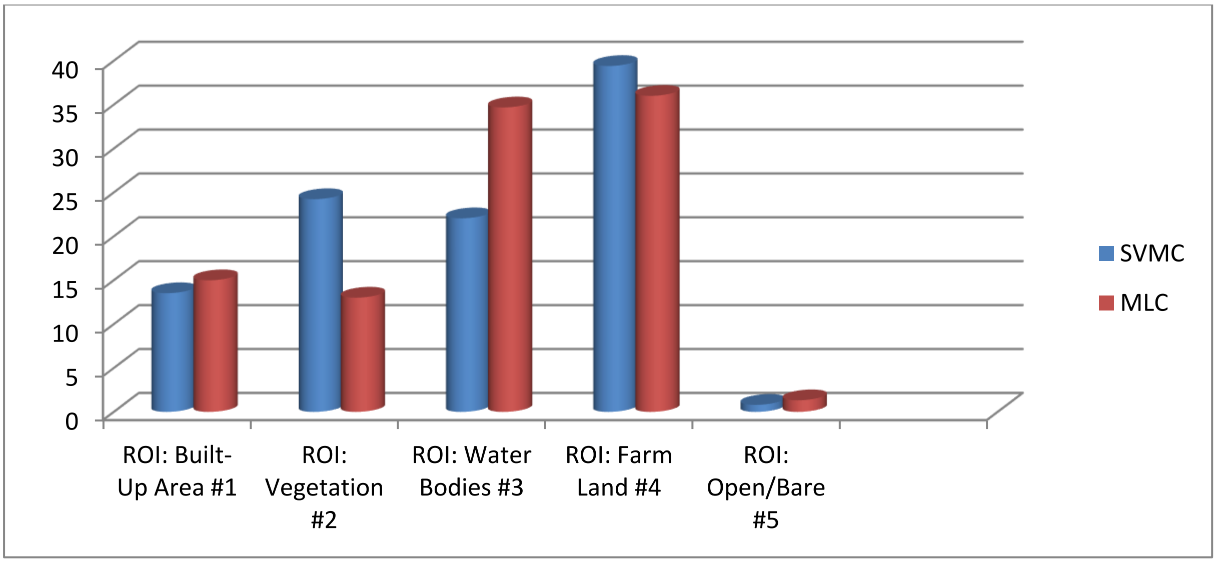


Figure 4.33: Land Use and Land Cover of Nnewi-North L.G.A, indicating SVM against ML

4.4. Procedures Adopted in Analysing Objective Number Four.

Objective No. 4 "To evaluate the performance of SVM and Maximum Likelihood (ML) in mapping geometric features using Error Matrix, Kappa and a newly proposed Post-Confusion Matrix (PoCoMa) template"was analyed as follows:

4.4.1. Post confusion matrix (statistical) analyses

This study observed that to further analyse the performance of SVM and Maximum Likelihood (ML) in mapping geometric features using only the overall accuracies and Kappa Statistical Procedures, will not satisfactory provide enough testing parameter required for Paired Samples Test or hypotheses testing and therefore created more testing parameters from their contingency matrix, kappa coefficient and the behavoural information of the classifiers during image classification. Thus, the project developed a "Post Confusion Matrix" (PoCoMa) template from the contingency matrix, kappa coefficient and the behavours of the image classifiers and used it for computing and comparing their capabilities of SVM and ML. The values obtained were tested using t-statistics (Paired Samples Test) using GraphPad Prism and SPSS statistical software.

The following procedure was considered in developing thePost-Confusion Matrix (PoCoMa) template.

i. Train your satellite image to obtain ROIs and Ground Truth Image. In this study, five classes of ROIs were adopted.

ii. Use the obtain ROIs (representative sample) and Ground Truth Image to perform image classification using any algorithm or classifier of your choice. In this study SVM and ML was used

iii. Perform error matrix and kappa coefficient analysis. This was done to obtain the following: User's Acuracy, Producer's Acurracy, Error of Omission, Error of Commission and Kappa Coefficient value.

iv. The results from (iii) above, error matrix and kappa coefficient analysis was recomputed using the Developed "Post Confusion Matrix" (PoCoMa) template, this is to obtain the mean average of the summation of the following: all Classes of Interest, User Acuracy, Producer's Acurracy, Error of Omission and Error of Commission

v. Incorporates the results from (iv) above, in "Post Confusion Matrix" (PoCoMa) table, with the conditions that may surround the image classification, for instance, type of classifier used (Support Vector Machine (SVM), Maximum Likelihood (ML), Neural Network (NNC), Bayesian Network (BA), K-Nearest Neigbour (KNN) among numerous others, Rule Base of the software used, the study used ENVI rule base for supervised classification., Time or Speed of accomplishing the task of classification, Divided by 3600, The memory space taken in the whole process of image analysis and classification and note also if the classification process was supervised or unsupervised classification? For example, if it is supervised, code it 1 and if it is unsupervised classification let the code be 0.

vi. Subject the whole parameters to statistical test. This is to compute the aggregate mean performance properties of classifiers used which must be morethan two in number. In this study, the PoCoMa template was used for computing and comparing their capabilities of two classifiers, SVM and ML. The values obtained were tested using t-statistics (Paired Samples Test) using GraphPad Prism and SPSS statistical software.

vii. Then accept or Reject the Null or Alternate Postulated statement, according to the outcome of the Hypotheses and

ix. Present your Results.

Table 4.9: Post Confusion Matrix (PoCoMa)Parameters

S/N	Parameters Code	Descriptions
1	OA	Overall Accuracy
2	KC	Kappa Coefficient
3	UA	User Accuracy
4	PA	Producer's Accuracy (Average Value)
5	EOO	Error of Omission (Average Value)
6	EOC	Error of Commission (Average Value)
7	SPD	Speed/Time (Divided By 3600)
8	SP	Space/Memory
9	RB	Rule Based
10	SV	Supervised
11	ROI	Region of Interests (FOIs)
12	COI	Classes of Interest (Average Value)

Table 4.10: Computation of the average performance of ML using Error Matrix result

	MAXIMUM LIKELIHOOD CLASSIFIER (ML)								
S/N	**PARAMETERS**	**BA**	**VG**	**WB**	**FL**	**OP**	**TOTAL**	**AVG.**	PARAMETERS
1	COI	14.99	13.01	34.08	36	1.32	99.4	**19.88**	Classes of Interest
2	UA	81.29	98.74	95.32	1.92	97.96	375.23	**75.05**	User Accuracy
3	PA	92.26	51.52	94.03	51.99	100	389.8	**77.96**	Producer's Accuracy
4	EOO	7.74	48.48	5.97	48.01	0	110.2	**22.04**	Error of Omission
5	EOC	18.71	1.26	4.68	98.08	2.04	124.77	**24.95**	Error of Commission
	Total	214.99	213.01	234.08	236	201.32	1099.4	**219.88**	
	Avg.	42.99	42.60	46.82	47.20	40.26	**219.87**		

From the designed parameters in (Table 4.9) the researcher developed a "Post Confusion Matrix" (PoCoMa) procedure that was used for computing and comparing the capabilities of different image classifier. The values obtained were tested using t-statistics (Paired Samples Test) where GraphPad Prism version 7.0 and SPSS version 21 statistical software was employed for the data analyses.

Table 4.11: Profileof the developed "Post Confusion Matrix" (PoCoMa) for ML

MLC DATA SET

S/N	PARAMETERS	ML	RECOMPUTED	DESCRIPTIONS
1	OA	82.5		Overall Accuracy
2	KC	0.76		Kappa Coefficient
3	UA	97.96	75.05	User Accuracy
4	PA	75.05	77.96	Producer's Accuracy
5	EOO	110.02	22.04	Error Of Omission
6	EOC	124.77	24.95	Error Of Commission
7	SPD	1020	0.283	Speed/Time(Divided By 3600)
8	SP	7.35		Space/Memory
9	RB	1		Rule Based
10	SV	1		Supervised
11	ROI	1		Region of Interests (FOIs)
12	COI		19.88	Avg. Classes Of Interest

Table 4.12: Computation of the average performance of SVM using Error Matrix result

SUPPORT VECTOR MACHINE CLASSIFIER (SVM)									
S/N	**CODE PARAMETERS**	**BA**	**VG**	**WB**	**FL**	**OP**	**TOTAL**	**AVG.**	**PARAMETERS**
1	COI	13.52	24.23	22.05	39.4	0.81	100.01	**20.02**	Classes of Interest
2	UA	99.83	89.76	98.5	99.15	100	487.24	**97.44**	User Accuracy
3	PA	98.5	96.45	99.64	91.1	100	485.69	**97.14**	Producer's Accuracy
4	EOO	1.5	3.55	0.36	8.9	0	14.31	**2.86**	Error of Omission
5	EOC	0.17	10.24	1.5	0.85	0	12.76	**2.55**	Error of Commission
	Total	213.52	224.23	222.05	239.4	200.81	1100.01	**220.01**	
	Avg.	42.70	44.85	44.41	47.88	40.16	**220.00**		

The essence of the "Post Confusion Matrix" (PoCoMa) shown in table 4.11 and 4.12 is to obtain the mean average of the summation of the following: all Classes of Interest,User Acuracy, Producer's Acurracy,Error of Omission andError of Commission, see table 4.13.

Table 4.13: Profileof the developed "Post Confusion Matrix" (PoCoMa) for SVM

SVM DATA SET				
S/N	**PARAMETERS**	**SVM**	**RECOMPUTED**	**DESCRIPTIONS**
1	OA	98.07		Overall Accuracy
2	KC	0.97		Kappa Coefficient
3	UA	117.448	97.44	User Accuracy
4	PA	117.138	97.14	Producer's Accuracy
5	EOO	14.31	2.86	Error of Omission
6	EOC	12.76	2.55	Error of Commission
7	SPD	172800	48	Speed/Time
8	SP	19.35		Space/Memory
9	RB	1		Rule Based
10	SV	1		Supervised
11	ROI	1		Region of Interests (FOIs)
12	COI		20.02	Classes of Interest

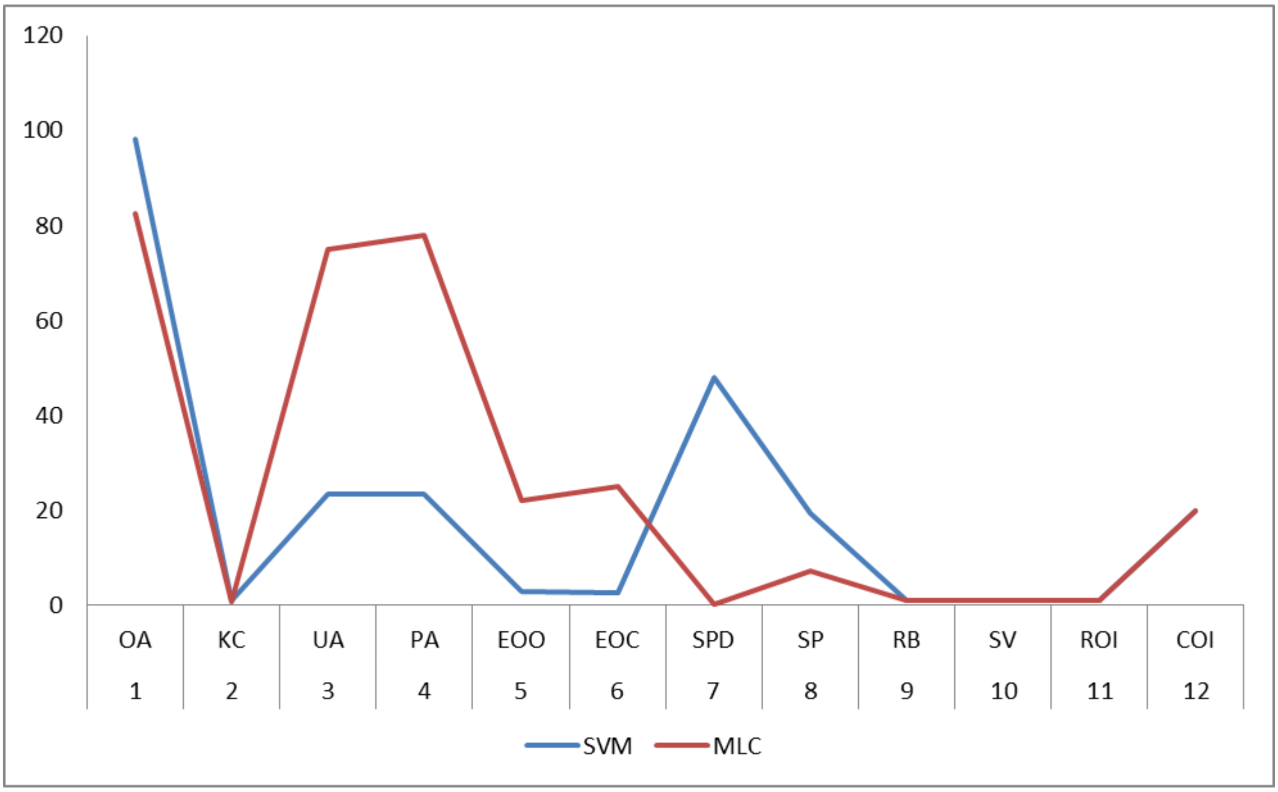


Figure 4.34: Diagrammatical display of the relationship between the performance properties of SVM and ML.

This relationship between the performance properties of SVM and ML was tested using Excel spreed sheet to first see if it can run in SPSS or GraphPad or not. If it contains error it may not run,Figure 4.34.

4.4.2. Hypotheses Testing

i. **Hypothesis One: The Null and Alternate Postulated statement tested is**

Null hypothesis (H_0): There is no significant difference between the performance of SVM and ML.

Table 4.14: Combine Dataset of Post Confusion Matrix (PoCoMa) from SVM and ML

S/N	CODE	SVM	ML	PARAMETERS
1	OA	98.07	82.5	Overall Accuracy
2	KC	0.97	0.76	Kappa Coefficient
3	UA	23.49	75.05	User Accuracy
4	PA	23.43	77.96	Producer's Accuracy
5	EOO	2.86	22.04	Error of Omission
6	EOC	2.55	24.95	Error of Commission
7	SPD	48	0.283	Speed/Time
8	SP	19.35	7.35	Space/Memory
9	RB	1	1	Rule Based
10	SV	1	1	Supervised
11	ROI	1	1	Region of Interests (FOIs)
12	COI	20.02	19.88	Avg. Classes of Interest

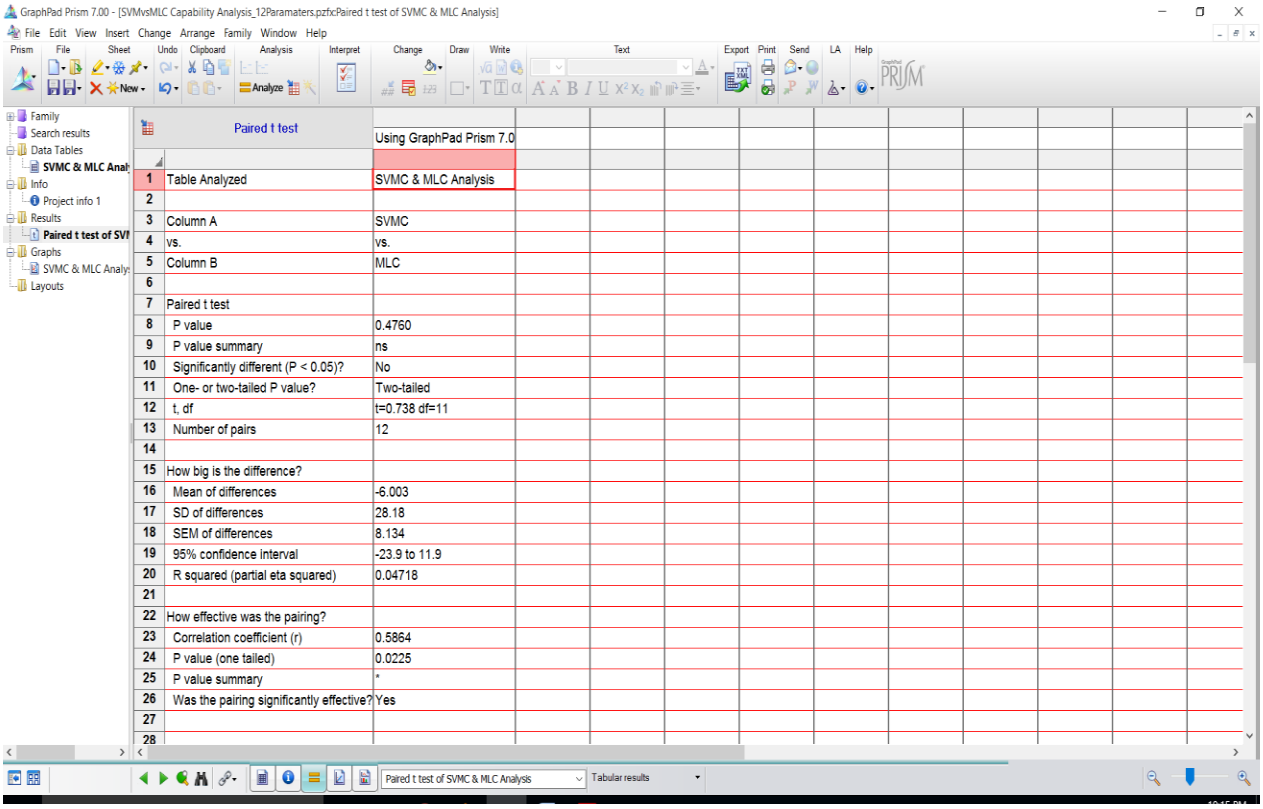


Figure 4.35: The summary sheet of the result of Hypotheses one in GraphPad Prism 7.0 window performance properties of SVM and ML

	Paired t test	
1	Table Analyzed	SVMC & MLC Analysis
2		
3	Column A	SVMC
4	vs.	vs.
5	Column B	MLC
6		
7	Paired t test	
8	P value	0.4760
9	P value summary	ns
10	Significantly different (P < 0.05)?	No
11	One- or two-tailed P value?	Two-tailed
12	t, df	t=0.738 df=11
13	Number of pairs	12
14		
15	How big is the difference?	
16	Mean of differences	-6.003
17	SD of differences	28.18
18	SEM of differences	8.134
19	95% confidence interval	-23.9 to 11.9
20	R squared (partial eta squared)	0.04718
21		
22	How effective was the pairing?	
23	Correlation coefficient (r)	0.5864
24	P value (one tailed)	0.0225
25	P value summary	*
26	Was the pairing significantly effective?	Yes

Figure 4.36: The extracted results of Hypotheses one analysis from GraphPad Prism 7.0 software

The single advantage of GraphPad Prism results(Figure 4.35) over the SPSS is that it is very easy to understand, even if you are not well grounded in statistical analysis and interpretations, for example, see the portion of Fig. 4.36 circled in red where “ns means not significant” that is, P Value isnot significant and ‘No’ there means that the test for hypothesis one has no significant different at P<0.05, moreover the whole result can be summarised and presented in a single tableas inFigure 4.35 and 4.36.

Table 4.15: The extracted results of Hypotheses one analysis fromSPSS version 21 statistical software

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	SVM	20.1450	12	28.51868	8.23263
	ML	26.1478	12	32.85397	9.48412

The result from table 4.15 and Fig.4.36 showed that the aggregate mean performance property of SVM is 20.15 while that of ML is 26.15. This tends to suggest that there could be a variation in the performance efficiency of map production and reading from SVM and MLC. However, the result from Fig.4.35, 4.36 and Table 4.16 compared the aggregate mean from SVM and ML to determine the significance of the mean difference. The result showed that the t-statistics is -0.738 with significant (probability) value of 0.476. Since the probability value is less than 0.05, we cannot reject the null hypothesis that "there is no significant difference in the result from SVM and ML". The study thus concludes that using any of the algorithms (SVM and ML) in ENVI yields no significance difference in performance and efficiency of result of the map produced. This implies that any of the SVM and ML is as good as each other. The result is also the same when tested using GraphPad Prism 7.0 software and SPSS version 21 statistical software.

Table 4.16: Comparative Analysis of the Aggregate Mean from SVM and ML usingSPSS

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 SVM - ML	-6.00275	28.17573	8.13363	-23.90476	11.89926	-.738	11	.476

Table 4.17: Paired Samples Correlations analysis using IBMSPSS

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 SVM & ML	12	.586	.045

4.5. Procedures Adopted in Analysing Objective Number Five.

Objective No. 5 "To produce the landuse and landcover map of Nnewi-North Local Government Area, Anambra State, Nigeria"was analysed as follows:

The landuse and land cover (LULC) that will serve as a base map of Nnewi North wasclassified using ENVI SVM and ML algorithms. After Image classification and Post-processing, including ground truthing/field verification and Computation of the error matrix to assess accuracy of the job, the result of image classification will be accepted if it is satisfactory, otherwise you may reject the result and retrain the image for better performance.

When this necessary condition has been achieved, the map is designed cartographically to show the necessary features of interest, result published in a digital format or printed as an analogue map. So, at this stage, the result of classified image in ENVI software was exported to ArcGIS 10.2 as vector layer for cartographical analysis and design. ENVI can easily convert the raster result to vector automatically by clicking at the icon (conversion to vector). Thereafter, the result of image classification was used for creating physical vector layers, that is, the polygon, line or point features layers of objects mapped. the following layers was created in Arc Catalog in a personal Geodatabase using the national town planning standard urban planning code.for instance, Commercial Area (Navy blue colour), Industrial Area (Pinkcolour), Educational (Schools) and Institutional/Administrative building (Redcolour), Residential Area or Mixed Landuse (Light Yellowcolour), Public Utilities (Browncolour), Recreational area which includes Green Spaces, Open Space & Parks (Light Greencolour), Areas Liable to Flooding (Light bluecolour), Agricultural area (Greencolour), other are Circulation/Roads (Blackcolour). Ifeanyi Uba Stadium, Local Government Area (LGA), Market (Nkwo Nnewi), Hospital (NAUTH) and Heritage site, among others were some of the prominent features captured.

Subsequently is the image digitizing using the already created feature layers and when it was completed another important thing here is inserting the following: the north direction, legend, scale (graphical and absolute statement scale), index, text, border line, grid lines and coordinates, title of the map among others were added to the data frame of the classified dataset to cartographically modified the proposed output map using the mentioned conventional symbols that can be easily understood globally by professional and unskilled people alike. The result of the map produced is shown in Fig4.37 and appendix B1 to B4.

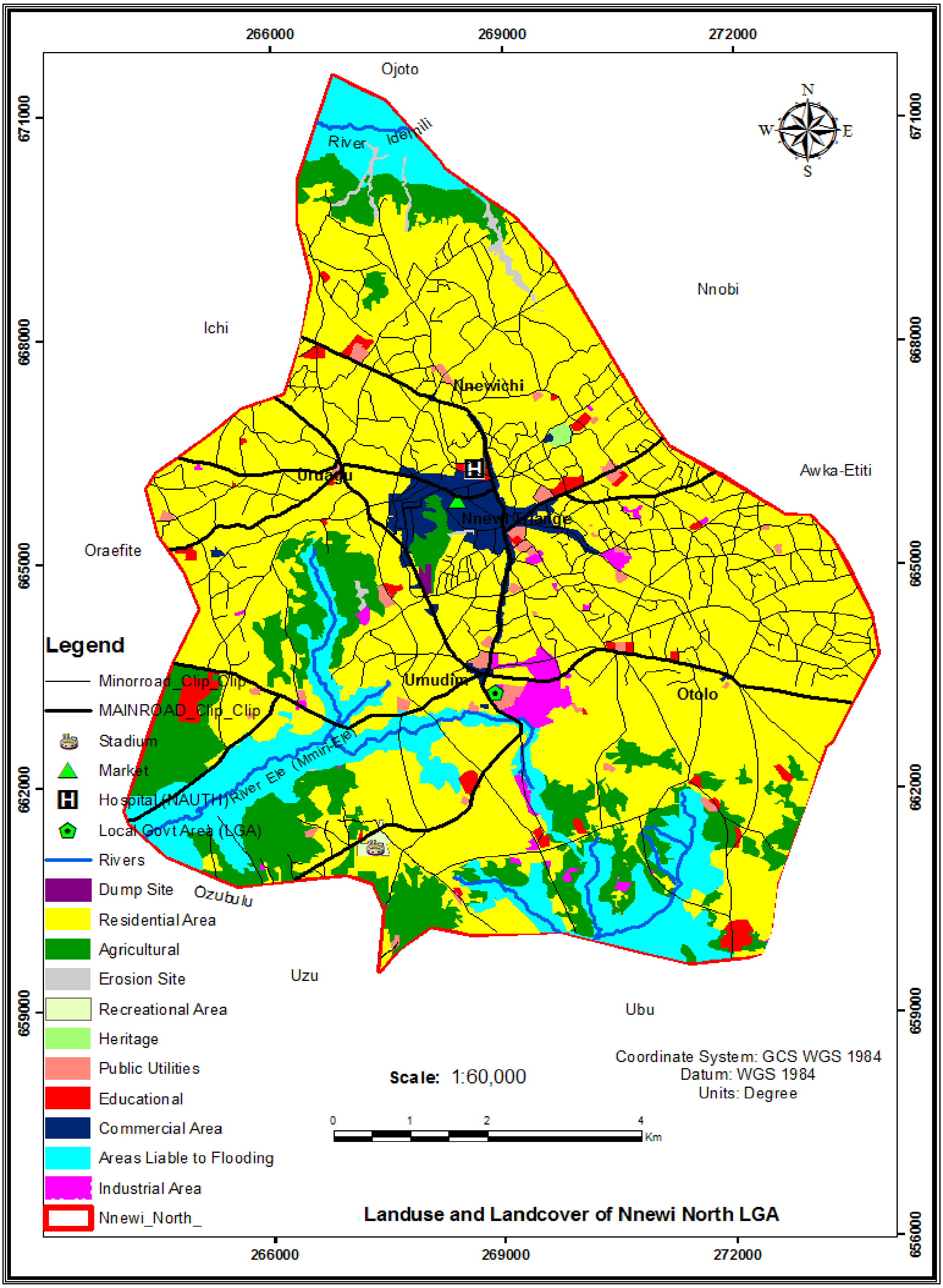


Figure 4.37: The physical vector map of Nnewi-North L.G. A.

The landuse and landcover map produced was converted into a physical vector map of Nnewi-North L.G.A using the planner's concept of urban delineation. Figure 4.37 was also printed in A3 paper and attached as an Appendix B4.

4.6Observations and inventory from the production of LULC map

The ENVI support Vector machine classifier (SVM) provides four types of kernels: linear, polynomial, radial basis function (RBF), and sigmoid. The default is the radial basis function kernel was used in the classification which works well in most cases. Classified satellite image of 2016 was used to produce the land use / land cover information of the study area, which revealed some salient question about the study area and also attempted answering them; for instance, there are questions like: (1). what is the settlement pattern of Nnewi-North LGA? 60% Nucleated settlement pattern was observed in the Central Business Distrect Area (CBDA) while the remanining 40% were dispersed settlement pattern observed immediately after theCBDA. (2). Does the city require urban renewal? Yes, because of the presence of some development without access road within the CBDA. Hence, urban renewal is required by widening the old roads, creating new ones and other basic infrastruture in the area to accomadate the unprecedented population increase and urbanization. (3). Does the city have land for future development? Yes in the sense that classified result says so. (4). The ownership of land determines what it is used for. The type of land ownership and title in Igbo land highly affects the way the state is being developed. Land is owned by individuals as a customary inheritance from their great grand fathers, hence, it may be difficult to convince one to lease it out for any reason. Rather they develop those lands by themselves without proper considerations. This was proven by the results from the researcher’s field observation and personal communication.

Plate 4.1: The researcher interviewing a respondent, at 'Nkwo-Nnewi' during the fieldwork.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1. Summary of findings

One of the justifications of this research is lack of adequate and correct geospatial information of Nnewi-North Local Government Area, Anambra State, Nigeria. The geospatial information will aid daily planning, development, monitoring and accurate decision-making by end users. Remember, since the history of States and Local government area creations in Nigeria, in 1967, 1976,1987, 1991and 1996 respectively, some of our administrative boundaries were not updated and one of them is the map of old Nnewi Local Government Area.The specimen of the map is illustrated in (Appendix A1)which is showing old Nnewi Local Government Area map comprising of some part of Ekwusigo LGA that are most commonly used for today administrative planning, development, monitoring and decision-making.Thus, this research sought and used subset of the satellite imagery covering Nnewi-North and South LGAin producing a baseline landuse and landcover map of study area using the correct shape file of Nnewi-North, (Appendix A2). To represent the correct land extent of Nnewi-North LGA after Ekwusigo LGA which was carved out of Nnewi-North in 1996 by General Sani Abacha (GCFR).

The contributions to knowledge and policy implications of the findings of the study are stated as follows;

5.1.1 Image 'Features of interest' (FOIs)extractionand their separability analysis was achieved

Features of Interest(FOIs) are portions of images, either selected graphically or selected by other means, such as thresholding. Then statistical analysis was targeted at achieving FOIs separability analysis were carried out with the Jeffries-Matusita and Transformed Divergence separability. The results of “Trained” and “Ground Truth” FOI indicate how well the selected FOI pairs are statistically separate, the values range from above 1 to greater than 1.9 which indicate that the FOI pairs have good separability.However, the summary of the separability index analysis shows that the Transformed Divergence (TD)performs better than the Jeffries-Matusita (JM) using both trained ROIs and Ground truth ROIs. Objective of this section of research was achieved and results demonstrated the best way of going about it and why it is important to validate the areas of classes the analyst trained before using them for supervised image classification.

5.1.2 Supervised classification using SVM and ML in ENVI Software

The ENVI SVM classifier provides four types of kernels: linear, polynomial, radial basis function (RBF), and sigmoid. The default which is the radial basis function kernel is used in this classification and which the developers and some researchers claimed works well in most cases. This study evaluated this claim by usingsupervised classification of remotely sensed data (GeoEye-1 HRSI) the result was compared using Maximum Likelihood (ML)in ENVI Software and the result shows that Support Vector Machine (SVM) performs better than theMaximum Likelihood (ML) in ENVI Software.However, it may be becauseSVMC has a way of dealing with complex and noisy data.The machine learning approaches improve the capability of classification in an intelligent way and hence the capability of SVM in ENVI software was confirmed using HRSI.

5.1.3Development and testing of ‘Post Confusion Matrix’ template was realized

The researcher observed that to further analyse the performance of SVM and ML in mapping geometric features using only the overall accuracies and Kappa Statistical Procedures will not satisfactory provide enough testing parameter required for Paired Samples Test or hypotheses testing and therefore created more testing parameters from their contingency matrix, kappa coefficient and the behavoural information of the classifiers during image classification. Thus, the project developed a "Post Confusion Matrix" (PoCoMa) template from the contingency matrix, kappa coefficient and the behavours of the image classifiers and used it for computing and comparing their capabilities of SVM and ML. The values obtained were tested using t-statistics (Paired Samples Test) using GraphPad Prism and SPSS statistical software.

The result fromt-statistics showed that the aggregate mean performance property of SVM is 20.15 while that of ML is 26.15. This tends to suggest that there could be a variation in the performance efficiency of map production and reading from SVM and ML. However, when the result was used to compared the aggregate mean from SVM and ML to determine the significance of the mean difference. The t-statistics result showed that the probability value is not less than 0.05, thus, the null hypothesis that "there is no significant difference in the result from SVM and ML"was not rejected.The study therefore concludes that using any of the Algorithms (SVM and ML) in ENVI yields no significance difference in performance and efficiency of result of the map produced. This implies that any of the SVM and ML is as good as each other when tested using GraphPad Prism 7.0 software and SPSS version 21 statistical software.

5.1.4. Production of landuse/land cover map of Nnewi-North was achieved

Remote sensing data, spectral reflectance extraction and statistical analysis of the data were applied. The percentage of built-up area shows that Nnewi still have available land for good developmental planning, that is, since out of 100%, its only 13.6% has been developed. Hence, urban renewal, proper zoning and coordination to have an adorable smart city should not be over emphasised. In this case, the production of the basemap of Nnewi-North L.G.A has been justified, which will be valuable for land-use planning and monitoring by end-users. Consequently, the GeoEye-1 Image of 2016 will serve as a base satellite image that will be used in change detection analysis in the near future that is to say, about 5 or 10 years it will serve as a yardstick for measuring urban dynamismof Nnewi-North L.G.A.

5.2 Conclusion

Creation and periodic updating of the base map of an urban area by assessing the built up area in relation to other land cover features through the input of experts in Surveying and Geoinformatics is a prerequisite for national development. The availability of various algorithms through machine learning approaches improves the capability of image classification in an intelligent way. Hence the capability of Support Vector Machine (SVM) algorithm incorporated in ENVI software was confirmed usingGeoEye-1, (2016) High Resolution Satellite Image (HRSI) to produce a reliable urban land cover map of Nnewi North Local Government Area, which will serve as a base map for land-use planning and monitoring by end-users.

The result showed that it is only the built-up areas and open/bare surfacesthat were well classified without conflicting with the spectral signatures of other urban features and this confirmed that SVM in ENVI software is good in detecting urban infrastructural developments. In addition, the result of the classification showed that Nnewi-North has enough land for future development, However, interview revealed that in Igbo culture, land is owned by individuals as a customary inheritance from their forefathers. This may consequently affect the way the state is being developed. Notwithstanding,the development of the city of Nnewi needs serious bottom-up planning rather than top-to-bottom planning approach in order to avoid any avoidable error looming around the city, like unplanned development, narrow roads/poor road network, poor traffic control, lack of open space and gardens among others which obviously occurs because of population growth, urbanization, and industrialization of the area. From existing information stored in the 1964 topographical map of Nnewi, the city was just a rural area with scattered family settlement pattern, population was less and amenities was little or none in existence. But now, the reverse is the case, thus, it is never too late to restructure the city for a sustainable urban development.

This study is hereby advocating for the integration of developmental activities in Nnewi North local government area into the basemap of the area, for instance, proper siting of new features at the right places. The application of the results and findings of this research is not only restricted to the study area, but extends to other urban areas located in Anambra state and beyond that bear similarity with the area of study.

5.3 Recommendations

The study recommends that

i. To better check the application of remote sensing in mapping urban features,FOIs separability analysis is veryimportant.
ii. The ENVI SVM is a robust tool for urban landscape mapping, especially for classifying man-made objects.
iii. Using the Basemap and Master Plan of the area will help the urban area to sustain healthy development and environment.
iv. There is serious need for proper land use policy implementation and monitoring of the daily development going on in Nnewi city for urban sustainability
v. The study reveals that Nnewi has rich agricultural lands and as such, recommend that urban agriculture be practice forsustainable urban livelihood and food security.
vi. There should be adequate physical planning that should take into consideration of the topography and features in the area so as to reduce the effect of current erosion, flooding and poor soil.

5.4. Further Studies:

From conclusions of the major outcomes of this study, the research areas that require ‘Further Studies’ are as follows;

a). Production of topographic map of Nnewi using HRSI: “that will serve as an applied input data to enable site selection and planning among other uses” will compliment this study.This is because, ‘3D Urban Mapping’ will enhance site design, hydrological analysis, suitability mapping and other benefits in delivering or achieving a smart city development and management.

b). Similar studies should be conducted to cover Nnewi-South Local Government Area and possibly other cities in Anambra State and beyond to aid site planning, land reclamations and urban renewal process.

c).Comparative analysis of the four Kernel types of SVM in ENVI software using HRSI and the developed “Post Confusion Matrix” formula.

d). Write the algorithm that will merged the existing ‘Error Matrix’ and the novel “Post Confusion Matrix” formulafor computing and comparing the capabilities of different image classifier.

f).Change Detection should be carryout to determine the urban dynamism, because the recent HRSI of 2016 reveals evidence of new development/buildings.

REFERENCES

Abe, Bolanle Tolulope (2014). Ensemble Classifiers for Land Cover Mapping. Unpublished PhD Thesis. The Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg.

Amanze J.O, Ezeh C.I and Okoronkwo, M.O (2015). Pattern of Income Diversification strategies among Rural Farmers in Nnewi North Local Government Area of Anambra. *Journal of Economics and Sustainable Development* 6 (5)

Amari,S.andWu,S., (1999). Improvingsupport vector machine classifiers by modifying kernel functions. *NeuralNetworks*, (12), 783-789.

Anambra State Ministry of Health and FHI 360. (2013). *Anambra State Wide Rapid Health Facility Assessment,* Nigeria: Anambra State Ministry of Health and FHI 360.

Anderson, J. R., Hardy, E.E., Roach, J.T., and Witmer, R.E (1976). A Land Use and Land Cover Classification System for Use with Remote Sensor Data. *A revision of the land use classification system as presented in U.S. Geological Survey Circular 671.Geological Survey Professional, 964.*

Anguita, D. Ridella S., Rivieccio F., Zunino R., (2003). Hyperparameter Design Criteria for Support Vector Classifiers. *Neurocomputing, Special Issue on Support Vector Machines.*

Anigbogu, T.U, Onwuteaka, C.I, Edoko, T.D, Okoli, M.I (2014). Roles of Small and Medium Scale Enterprises in Community Development: Evidence from Anambra South Senatorial Zone, Anambra State.*International Journal of Academic Research in Business and Social Sciences, 4(8,)*

Anis Charrada and Abdelaziz Samet (2012). Complex Support Vector Machine Regression for Robust Channel Estimation in LTE Downlink System. *International Journal of Computer Networks & Communications* (IJCNC) 4(1),1-14.

AnujaKumari V. and Chitra, R. (2013). Classification of Diabetes Disease Using Support Vector Machine. *International Journal of Engineering Research and Applications* (IJERA) 3(2), 1797-1801.

Arun, P. V. and Katiyarb S.K. (2013). An Intelligent Approach towards Automatic Shape Modelling and Object Extraction from Satellite Images Using Cellular Automata Based Algorithm. *Journal of GIS Science & Remote Sensing* 50(3), 337-348.

ASPRS, (2004): Manual of Photogrammetry, 5th Ed. ASPRS.

Aymerich Ismael F., PieraJaume, Mohr Johannes, Aureli Soria-Frisch and Klaus Obermayer. (2010). Potential Support Vector Machines for phytoplankton fluorescence spectra classification: Comparison with Self-Organizing Maps. Instrumentation view point, https://upcommons.upc.edu/bitstream/handle/2099/8926/Potential%20support%20vector%20machines%20for%20phytoplankton%20fluorescense%20spectra%20classification.pdf?sequence=1&isAllowed=y

Aymerich, I. F., Piera, J. and Soria-Frisch, (undated).A rapid technique for classifying phytoplankton fluorescence spectra based on self-organizing maps. *Applied Spectroscopy* (in press).

Baltsavias, E.P., (2004). Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58 (3-4), 129– 151.

Bayoudh Meriam, Roux Emmanuel, Nock Richard, and Gilles Richard (2012). Automatic Learning of Structural Knowledge from Geographic Information for updating Land Cover Maps. *Symposium of the Latin American Society for Remote Sensing and Spatial Information Systems (SELPER) Cayenne, French Guiana.IRD*, 10 p.

Bekkari Aissam, Soufian eIdbraim, Azeddine Elhassouny, Driss Mammass, Mostafa El yassa and Danielle Ducrot. (2012). SVM classification of high resolution urban satellites Images using Haralick features. *International Journal of Scientific & Engineering Research*, 3(6),

Bellman Chris J. and Shortis Mark R. (2014). Using Support Vector Machines for Building Recognition in Aerial Photographs. Commission III, Working Group III/4 https://www.researchgate.net/publication/228920054_Using_support_vector_machines_for_building_recognition_in_aerial_photographs. Available online (Date retrieved 10/23/2014).

Blaschke, T., Lang, S., (2006). Object based analysis for automated information extraction-a synthesis. In: MAPPS/ASPRS Fall Conference, San Antonio, TX.

Blaschke, T., (2003). Object-based contextual image classification built on image segmentation, in *Proc. IEEE Workshop Adv. Tech. Anal. Remote Sensing*, pp.113-119.

Bohara, A., Mitchell, N., Mittendorff, C. (2004). Compound democracy and the control of corruption: A cross-country investigation. The Policy Studies Journal 32(4) 481–499

Boley Daniel and Caoy Dong Wei (2004). Training Support Vector Machine using Adaptive Clustering. *A proceeding of 2004 SIAM International Conference on Data Mining*, April 22 - April 24, Orlando, FL, USA.

Bortolot, Z. (1999). "Image classification", Online publication http://www.geog.ubc.ca/courses/geog570/notes/classification.html. (retrieved 10/23/2014).

Boser, B., Guyon, I., and Vapnik, V. (1992). A Training Algorithm for Optimal Margin Classifiers. *In proc. of 5th ACM Annual Workshop on Computational Learning Theory, Pittsburgh, Pennsylvania*, pp. 144-152.

Braimoh, A.K., &Onishi, T. (2007). Spatial determinants of urban landuse change in Lagos, Nigeria. *LandUse Policy*, *24*(2), 502-515.

Brautigam, D. (1997). Substituting for the State: Institutions and Industrial Development in Eastern Nigeria. *World Development* 25(7), 1063-1080.

Brenner, C., (2005). Building reconstruction from images and laser scanning, International *Journal of Applied Earth Observation and Geoinformation*, 6 (3-4), 187–198.

Bruzzone, L., Carlin, L., (2006). A multilevel context-based system for classification of very high spatial resolution images. IEEE Trans. *Geosci. Remote Sens.* 44 (9), 2587–2600. doi:10.1109/TGRS.2006.875360.

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, (2), 121–167.

Burnett C. and Blaschke, T. (2003). A multi-scale segmentation/object relationship modeling methodology for landscape analysis. *Int. J. Ecol. Model. Syst. Ecol*., 168(3), 233–249.

Butler, W. L. and Hopkins, D. W. (1970). Higher Derivative Analysis of Complex Absorption Spectra. *Photochem. Photobiol*. 12, 439-450.

Caridade CMR, Marçal ARS, Mendonça T, (2008). The use of texture for image classification of black and white air-photographs. Int J Remote Sens. 29(2),593-607. doi:10.1080/01431160701281015.

Černá, Lenka & Chytrý, Milan, (2005). Supervised classification of plant communitieswith artificial neural networks. Journal of Vegetation Science 16, 407-414.

Chang, C.C. and Lin, C. J. (2001): LIBSVM: a library for support vector machines. Chang, C-I, Wu, C-C., Liu, W-m. and Ouyang, Y-C. (2006). "A new growing method for simplex-based endmember extraction algorithm", IEEE Trans. Geosci. & Remote Sensing, vol. 44, no. 10, pp. 2804-2819.

Chaudhry, F., Wu, C., Liu, W., Chang, C-I and Plaza, A. (2006). "Pixel purity index- based algorithms for endmember extraction from hyperspectral imagery", In Recent Advances in Hyperspectral Signal and Image Processing", C.-I Chang, Ed. Trivandrum, India: Research Signpost, vol. 3, pp. 31-61.http://www.csie.ntu.edu.tw/~cjlin/libsvm. (retrieved 10/23/2014).

Chanussot, J. Fauvel, M. Gamba, P. Gualtieri, A. Marconcini, M. Tilton, J.C. and TriannI, G. (2009). Recent Advances in Techniques for Hyperspectral Image Processing. *Remote Sensing Environment*, 113 (1), 110-122.

Christodoulos Psaltis, Charalabos Ioannidis, (2010). Supervised Change Detection on Simulated Data Employing Support Vector Machines. TS 7D - LIDAR and InSAR Usage in Surveying. *A proceeding of FIG Congress 2010, Facing the Challenges – Building the Capacity. Sydney, Australia*, 11-16 April 2010.

CIA Factbook: (http://www.cia.gov/cia/publications/factbook/).(retrieved 10/23/2014).

Cihlar, J., Xiao, Q., Chen, J., Beaubien, J., Fung, K. and Latifovic, R. (1998). "Classification by progressive generalization: A new automated methodology for remote sensing multichannel data. International Journal of Remote Sensing, 19(14), 2685 – 2704.

Clarke, K.C., (2001). Getting Started with Geographic Information Systems, Upper Saddle River: Prentice-Hall, 353.

Cohen L, Manion L, Morrison K. (2000). Research Methods in Education. 5th ed. London: RoutledgeFalmer.

Colwell, R.N (Ed.), (1983). *Manual of Remote Sensing*, 2nd Ed., Falls Church: ASP&RS.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment 37(1), 35-46.

Corina, I., Didier, B., Matthieu, C., (2008). Detection, characterization, and modeling vegetation in urban areas from high-resolution aerial imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1(3), 206-213.

Corruption: A cross-country investigation. *The Policy Studies Journal* 32(4) (2004).
Cortes Corinna and Vapnik Vladimir (1995). Support-Vector Networks. *Machine Learning*. 20(3), 273-297.

Cracknell, A.P and Hayes L.W.B. (1993). *Introduction to Remote Sensing,* London: Taylor and Francis, 293.

Cristianini N., and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press.

Curran, P.J., (1987). Remote Sensing Methodologies and Geography. *Intl. Journal of Remote Sensing. (*8), 1255-1275.

Dahlberg. R.W and Jensen, J.R (1986). Education for Cartography and Remote Sensing in the service of an Information Society: The United States Case. *American Cartographer,* 13 (1), 51-71.

Davis, B.A. (1999). Overview of NASA's Commercial Remote Sensing Program. *Earth Observation,* 8(3), 58-60.

Decherchi Sergio, Ridella Sandro, and Zunino Rodolfo, (2009). Using Unsupervised Analysis to Constrain Generalization Bounds for Support Vector Classifiers, 473. R1.ieeexplore.ieee.org/iel5/72/5420315/05401065.pdfdigilander.libero.it/sedekfx/papers_/vqsvm.pdf.

Dehqanzada, Y.A. and Florini A.M (2000). *Secrets for Sale: How Commercial Satellite Imagery Will Change the World.* Washington: Carneggie Endowment for intl. Peace, 45.

Delve (1996). Data for Evaluating Learning in Valid Experiments. http://www.cs.toronto.edu/~delve.(Available online, retrieved 10/21/2015).

Do Hiep-Thuan, Nguyen-Khang Pham, Thanh-Nghi Do (2005). A Simple, Fast Support Vector Machine Algorithm for Data Mining". Fundamental *&Applied IT Research Symposium* 2005.

Do, T-N, and Poulet, F. (2005). Mining Very Large Datasets with SVM and Visualization". in proc. of ICEIS'05, *7th Int. Conf. on Enterprise Information Systems,* 2, 127-134, Miami, USA.

Dongarra, J., Pozo, R., and Walker, D. (1993). LAPACK++: a design overview of object-oriented extensions for high performance linear algebra. *In proc. of Supercomputing'93, IEEE Press,* pp. 162-171.

Duan, T. D., Duc, D. A., Le, T. & Du, H. (2004). Combining Hough transform and contour algorithm for detecting vehicle license plate.*Proc. of International Symposium Intelligent Multimedia, Video and Speech Processing*, pp.747-750.
Durgesh K. Srivastava, and Lekha Bhambhu, (2009). Data Classification Using Support Vector Machine. *Journal of Theoretical and Applied Information Technology,* © 2005 - 2009 JATIT.

Dwivedi, R.S., Kandrika, S. and Ramana, K.V. (2004). Comparison of classifiers of remote-sensing data for land use/land cover mapping. Current Science, 86 (2), 328-335.

Eastman, R.J. (2006). Guide to GIS and Image processing. Clark University Worcester, pp. 87 –131.

Eneche, P.S.U, Ahmed, A. & Ekebuike, A. N. (2018). Application of VIS and Land Change Modelers in Characterizing and Predicting the Urban Land Use/Cover Dynamics of Nnewi Metropolis, Nigeria. *Selected Works of Confluence Journal of Environmental Studies (CJES), Kogi State University, Nigeria, 12 (1), 63- 79.*

ERDAS Field Guide (1999). University of Massachusetts, Amherst, MA, USA. Erdas Inc., Atlanta, Georgia. *Technical Report 92-8.*

Erener, A and Düzgün, H.S. (2009). A methodology for land use change detection of high resolution pan images based on texture analysis" Italian Journal of Remote Sensing - 2009, 41(2), 47-59.

FAO, (2005). Land Cover Classification System (LCCS) Classification Concepts and User's Manual. FAO, Rome, Italy.

Fauvel, M., Chanussot, J., Benediktsson, J.A., (2009). Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas. *EURASIP Journal on Advances in Signal Processing Article* ID 783194.

Felzenszwalb, P.F, McAllester, D. and Ramanan, D. (2008): A discriminatively trained, multiscale, deformable part model. In CVPR.

Fisher, P.F. and Lindenberg, R.E. (1989). On Distinctions among Cartography, Remote Sensing and Geographic Information Systems" *Photogrametric Engineering and Remote Sensing* 55(10), 1431-1434.

Foody, G.M., (1986). Approaches for the production and evaluation of fuzzy land cover classification from remotely sensed data. International Journal of Remote Sensing 17, 1317–1340.

Foody, G.M., and Mathur, A., (2004a). A relative evaluation of multiclass image classification by support vector machines". IEEE *Transactions on Geoscience and Remote Sensing* 42 (6), 1335–1343.

Foody, G.M., and Mathur, A., (2004b). Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment* 93 (1–2), 107–117.

Franke, J. Roberts, DA, Halligan, K, Menz, G (2009). Hierarchical multiple endmember spectral mixture analysis (MESMA) of hyperspectral imagery for urban environments. Remote Sens Environ 113, 1712–1723.

Franklin, S.E., M.A. Wulder, and M.B. Lavigne, (1996). Automated derivation of geographic window sizes for remote sensing digital image texture analysis. *Computers and Geosciences*, 22, 665–673.

Freedom House (2005): Freedom in the world country ratings.

Fukunaga, K. (1990). Introduction to Statistical Pattern Recognition, 2nd Ed. New York: Academic Press.

Fung, T. & Ledrew, E. (1988). The Determination of optimal threshold levels for change detection using various accuracy indices. Photogrammetric Engineering and Remote sensing Journal 54 (10), 1449 – 1454.

Fussell, J., Rundquist, D. and Harrington, J.A (1986). On Defining Remote Sensing. *Photogrammetric Engineering and Remote Sensing*, 52 (9), 1507-1711.

Geneletti, D., Gorte, B.G.H., (2003). A method for object oriented land cover classification combining landsat TM and aerial photographs. International Journal of Remote Sensing 24, 1273–1286.

Ghaffarian, Salar and Ghaffarian Saman (2014). Automatic Building Detection Based On Supervised Classification Using High Resolution Google Earth Images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Volume XL-3, 2014 ISPRS Technical Commission III Symposium, 5 – 7 September 2014, Zurich, Switzerland

Gong P, Howarth PJ. (1990). Frequency-based contextual classification and grey level vector reduction for land-use identification. Photogramm Eng Rem S. 1992 (58), 423-437.

Gong, P. and Zhang, A. (1999). Noise effect on linear spectral unmixing", Geographic Information Sciences, 5 (1), 52–57.

Graña M., and D'Anjou A. (2004) Feature Extraction by Linear Spectral Unmixing. In: Negoita M.G., Howlett R.J., Jain L.C. (eds) Knowledge-Based Intelligent Information and Engineering Systems. KES 2004. Lecture Notes in Computer Science, vol 3213. Springer, Berlin, Heidelberg doi:10.1007/978-3-540-30132-5_95.

Green, R. O., Eastwood, M. L., Sarture, C. M., Chrien, T. G., Aronsson, M., Chippendale, B. J., Faust, J. A., Pavri, B. E., Chovit, C. J., Solis, M. S., Olah, M. R. and Williams, O. (1998). Imaging spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). Remote Sens. Environ., 65(3), 227–248. doi: 10.1016/S0034-4257 (98)00064-9.

Gunn, R.S. (1998). Support Vector Machines for classification and regression, "t*echnical report*" University of Southampton.

Guo, Q., Kelly, M., and Graham, C. H., (2005), Support vector machines for predicting distribution of sudden oak death in California. *Ecological Modeling*,182,75-90.

Guyon, I.and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157–1182.

Haala, N., Kada, M., (2010). An update on automatic 3D building reconstruction. ISPRS *Journal of Photogrammetry and Remote Sensing*, 65(6), 570–580.

Han, Jiawei and Kamber, Micheline (2012). Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, USA, 3rd Edition.

Haralick R. M, (1979). Statistical and structural approaches to texture. *Proceedings of IEEE*; 67:786-804. doi: 10.1109/PROC.1979.11328.

Harralick R. M., Shanmugam K., and Dinstein I., (1973). Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, 3(6), 610–621.

Hastie,T.,andTibshirani,R., (1998).Classificationbypairwise coupling.InJordan,M.I., Kearns, M.J., Solla, A.S. (eds.),*Advancesin NeuralInformationProcessingSystems*,Vol.10, MIT Press, Cambridge.

Hay, G., Blaschke, T.., (2003). A comparison of three image-object methods for the multiscale analysis of the landscape structure. *ISPRS Journal of Photogrammetry and Remote Sensing* 57, 327–345.

Haykin, S. (1999). Neural Networks: A Comprehensive Foundations, 2ed. Upper Saddle River: Prentice Hall.

Hecher, J. Filippi, A. Guneralp, I. and Paulus, G. (2013). Extracting River Features from Remotely Sensed Data: An Evaluation of Thematic Correctness. Doi: 10.1553/giscience 2013s187.

Heiden U, Segl K, Roessner S, Kaufmann, H (2007). Determination of robust spectral features for identification of urban surface materials in hyperspectral remote sensing data. *Remote Sens Environ* 111,537–552.

Heinz, D.C. and Chang, C. (2001). Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Transactions Geoscience & Remote Sensing*, 39 (3), 529-545.

Herold, M., Gardner, M.E., & Roberts, D.A. (2003). Spectral resolution requirements for mapping urban areas. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 1907-1919.

Hochreiter S., and Obermayer, K. (2006). Support vector machines for dyadic data. Neural Computation, 18 (6), 1472-1510.

Hsu Chih-Wei and Lin Chih-Jen (2002). A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Transactions on Neural Networks*, 13 (2),

Hu Wenjie, Song Qing (2004). An accelerated decomposition algorithm for robust support vector machines. *IEEE Transactions on Circuits and Systems II: Express Briefs.* 51(5), 234-240.

Huang,C.,Davis,L.S.,andTownshed,J.R.G., (2002).Anassessmentofsupportvector machinesforlandcoverclassification.*InternationalJournalofRemoteSensing*,23(4), 725-749.

Huang, C., Song, K., Kim, S., Townshend, J. R. G., Davis, P., Masek, J. G and. Goward S. N. (2008). Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sensing of Environment*, 112(3), 970-985.

Huang, Chien-Jen (2011). Using Genetic Algorithm Optimization SVM to Construction of Investment Model. *International Journal of Digital Content Technology and its Applications*, 5(1).

Hung, et al. (2008) "Feature selection and classification model construction on type 2 diabetic patient's data. *Journal of Artificial Intelligence in Medicine,* Elsevier, 251-262.

Huysmans, Johan Martens, David Baesens, Vanthienen1, Bart Jan and Gestel Tony Van (2006). Country Corruption Analysis with Self Organizing Maps and Support Vector Machines. *Conference Proceedings on the Intelligence and Security*

Informatics, International Workshop, WISI 2006, Singapore, April 9, 2006, doi: 10.1007/11734628_13.

Hyvarynën, A., Karhunen, J., and Oja, E. (2001) Independent Component Analysis. John Wiley & Sons, New York.

Igbokwe (2006). Mapping and Spatial Characterization of Major Urban Centres in Parts of South Eastern Nigeria with Nigeriasat-1 Imagery. *A proceeding of the 5th FIG Regional Conference Accra, Ghana, March 8-11, 2006 on Promoting Land Administration and Good Governance.*

Iovan Corina, Boldo Dider, and Cord Matthieu (2008): "Detection, Characterization and Modelling Vegetation in Urban Areas from High Resolution Aerial Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(20).

Iovan Corina, Boldo Dider, and Cord Matthieu. (2014). Detection, Characterization and Modelling Vegetation in Urban Areas from High Resolution Aerial Imagery. *European Journal of Remote Sensing* 47: 717-738

Israel, S. and Kasabov, N., (1997). Statistical, connectionist and fuzzy inference techniques for image classification. *Journal of Electronic Imaging*, 6(3), 1-11.

Jensen, T. R., (2005). *Introductory Digital Image Processing:* A *Remote Sensing Perspective*, Pp.525. Prentice-Hall, Upper Saddle River.

Jensen, T. R., (2007). *Remote Sensing of the Environment: An Earth Resource Perspective*. Pearson Prentice Hall, United States of America. 2nd Ed. Chapter one.

Jeyanthi, S.N., (2007). *Efficient Classification Algorithms using SVMs for Large Datasets* (Master's thesis). Supercomputer Education and Research Center, IISC, Bangalore, India.

Jin, Hang and Feng, Yanming and Shen, Yan (2011). Accurate urban road model reconstruction from high resolution remotely sensed imagery based on Support Vector Machine and Gabor filters. *Proceedings of Joint Urban Remote Sensing Event (JURSE 2011*), 11-13 April 2011, Munich, Germany.

Joachims, T., (1998). Making large-scale SVM learning practical. In: B. Scholkopf, Burges, C.J.&Smola, A, J. (Eds.), *Advances in kernel methods - support vector learning.* MIT Press, Cambridge, USA.

Joachims, T., (1999). Estimating the Generalization Performance of an SVM Efficiently. In *ProceedingsofICML-00,17thInternationalConferenceonMachineLearning*, Morgan Kaufmann Publishers, San Francisco.

John Pike (2006). "ORBIMAGE Completes Acquisition of Space Imaging; Changes Brand Name to GeoEye" (http://www.globalsecurity.org/intell/library/news/2006/intell-060112-geoeye.htm). Globalsecurity.org. Retrieved 17/2/2015.

Jorgensen, J. (2004). "Remote sensing", Online publication: *http://maps.unomaha.edu/Peterson/GIS/notes/RS2.htm*, Cited 9 July2011.

Jungerius, P.O, Badwen MG, Obihara C.H (1964). "*Anambra-Do river area" Soil Survey Memoir*, No.1. Govt. Printer, Enugu. Nig.

Kapoor A., Mehta N., Esper F., Poljsak-Prijatelj M., Quan P. L., Qaisar N., Delwart E., Lipkin W. I. (2010). Identification and characterization of a new bocavirus species in gorillas. PLoS ONE 5, e11948 10.1371/journal.pone.

Keshava, N., and Mustard, J.F. (2002) Spectral unimixing. IEEE Signal Proc. Mag. 19(1), 44–57.

Keuchel,J., Naumann, S., Heiler, M., Siegmund A. (2003), Automatic land cover analysis for Tenerife by supervised classification using remotely sensed data. Remote Sensing of Environment 86(4):530-541 DOI: 10.1016/S0034-4257(03)00130-5.

Khoshelham, K., Nardinocchi, C., Frontoni, E., Mancini, A., and Zingaretti P. (2010). Performance evaluation of automated approaches to building detection inmulti-source aerial data. *ISPRS Journal of Photogrammetry and Remote Sensing* 65 123_133 www.elsevier.com/locate/isprsjprs.

Kidd, E.: XML-RPC How to. 2001. http://xmlrpc-c.sourceforge.net/xmlrpchowto/-xmlrpchowto.html Available online, retrieved 10/21/2015.

Kim, H., Pang, S., Je, H., Kim, D. and Bang, S.Y., (2003), Constructing support vector machine ensemble. Pattern Recognition, 36, pp. 2757–2767.

Knerr, S., Personnaz, L., Dreyfus G. (1990): *Single-Layer Learning Revisited*: A Stepwise Procedure for Building and Training a Neural Network. In Neuro-computing: Algorithms, Architectures and Applications, Fogelman-Soulie and Herault (eds.). NATO ASI Series, Springer, 41-50.

Koetza B., Morsdorfa F., Curtb T., van der Lindenc S., Borgnietb L., Odermatta D., Alleaumeb S., Lampinb C., Jappiotb M. and Allgöwerd B. (2006). Fusion of Imaging Spectrometer and Lidar Data Using Support Vector Machines for Land Cover Classification in the Context of Forest Fire Management.http://www.isprs.org/proceedings/XXXVI/7-C50/papers/P29.pdf.

Kramer J. H., (2002). *Observation of the earth and its environment*: Survey of missions and sensors (4th Ed). Berlin: Springer.

Kwesi Nooni Isaac (2012). *Oil Palm Mapping Using Support Vector Machine with Landsat ETM+ Data. (*Master's Thesis Geo-Information Science and Earth Observation Enschede, the Netherlands.

L'uborLadický, Philip.Torr H.S and Zisserman Andrew (2012). Latent SVMs for Human Detection with a Locally Affine Deformation Field. https://www.inf.ethz.ch/personal/ladickyl/ladf_bmvc12.pdf.

Landgrebe, D. (1998). The evolution of landsat data analysis. *Photogrammetric Engineering and Remote Sensing*, 63 (7), 859 – 867.

Landgrebe, D. (2005). Multispectral land sensing: Where from, where to? *IEEE Transactions on Geoscience and Remote Sensing*, 43, 433 - 440.

Lawrence, R., Bunn, A., Powell, S. and Zmabon, M (2004). Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing of the Environment* 90, 331–336.

Lee, Y. J., and Mangasarian, O. L., (2000). RSVM: reduced support vector machines. *Technical Report* 00-07, Data Mining Institute, Computer Sciences Department, University of Wisconsin-Madison, Wisconsin.

LevinN.(1999).“Fundamentalsofremotesensing”,a bookcompiledusingInternet resources, articlesand personal knowledge, 225.

Li Maokuan, Cheng Yusheng, Zhao Honghai (2004): "Unlabeled data classification via SVM and k-means Clustering". *Proceedings of the International Conference on Computer Graphics, Image and Visualization (CGIV04), IEEE.*

Li, R., Wang, W. and Tseng, H.-Z., (1998): "Object recognition and measurement from mobile mapping image sequences using hopfiled neural networks": Part 1, *ASPRS Annual Conference. American Society of Photogrammetry and Remote Sensing, Tampa, Florida, USA.*

Lillesand, T.M. & Kiefer, R.W. (1987). *Remote sensing and image interpretation*. John Wesley, & Sons, New York, 2nd Edition.

Lillesand, T.M., Kiefer, R.W., 1999. Remote Sensing and Image Interpretation. John Wiley and Sons Ltd., New York.

Lin Chun-Fu and Wang Sheng-De (2002). Fuzzy support vector machines. *IEEE Transactions on Neural Networks,* 13(2), 464-471.

Lin Chun-Fu and Wang Sheng-De (2003). Training algorithms for fuzzy support vector machines with noisy data. *IEEE XI1 Workshop on Neural Networks for Signal Processing,* 517-526.

Lin, Chun-Fu and Wang Sheng-De (2004). Training algorithms for fuzzy support vector machines with noisy data. *Pattern Recognition Letters,* 25, 1647-1656.

Liu Junping, Chang Mingqi, and Ma Xiaoyan, (2009): Groundwater Quality Assessment Based on Support Vector Machine. Funded by Global Environment Fund (GEF) Integral Water Resource and Environment Management of Haihe River basin (MWR-9-2-1),"111" Introducing Intelligence Project (B08039). http://www.seiofbluemountain.com/upload/product/201005/2009shzyhy03a1.pdf.Available online (retrieved 10/23/2014)

Liu, C. Yuen, J, Torralba, A, Sivic, J., and Freeman W. T. (2008): Sift flow: Dense correspondence across different scenes. In ECCV, 2008.

Liu,H.,Fan,Y.,Deng,X.and Ji, S.(2009). ParallelProcessing Architectureof Remotely SensedImage ProcessingSystemBasedonCluster",*Image and SignalProcessing,9, CISP'09. 2nd International Congress on*, 1-4.

Loung, G. and Tan, Z., (1992). Stereo matching using artificial neural networks.

Lu D. and Weng Q. (2006). A survey of image classification methods and techniques for improving classification performance, Department of Geography, Geology, and Anthropology, Indiana State University, Terre Haute, IN 47809, USA.

Lu, D. and Weng, Q. (2005). Urban land-use and land-cover mapping using the full spectral information of Landsat ETM+ data in Indianapolis, Indiana. *Photogrammetric Engineering & Remote Sensing,* 71(11): 1275-1284

Lu, D. and Weng, Q. (2006). Use of impervious surface in urban land use classification. *Remote Sensing of Environment*, 102(1-2): 146-160.

Lu, D., Weng, Q., (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing* 26 (5), 823–870.

Lucieer, V., (2008). Object-oriented classifications of side scan sonar for mapping benthic marine. *International Journal of Remote Sensing* 29 (3), 905–921.

Mahesh, P., and Mather, P.M., (2003). Assessments of the effectiveness of the decision tree method for land cover classification. *Remote Sensing of the Environment* 86, 554–565.

Mahi, Habib, Hadria Isabaten, and Chahira Serief, (2014). Zernike Moments and SVM for Shape Classification in Very High Resolution Satellite Images. *The International Arab Journal of Information Technology*, 11, (1).

Marceau, D.J., P.J. Howarth, J.M. Dubois, and D.J. Gratton, (1990). Evaluation of the grey-level co-occurrence matrix method for land cover classification using SPOT imagery, *IEEE 7 Transactions on Geoscience and Remote Sensing*, 28:513-517.

Martinez, P.J., Perez, R. M., Plaza, A., Aguilar, P.L., Cantero, M.C. and Plaza, J. (2006). Endmember extraction algorithms for hyperspectral images, *Annals of Geophysics*, 49, (1), 93-101.

Martino, S., Falanga, M., and Godano, C.: (2004). Dynamical similarity of explosions at Stromboli volcano, Geophys. J. I., 157, (3), 1247–1254, doi:10.1111/j.1365-246X.2004.02263. x.

Mather, P.M., (2004). *Computer Processing of Remotely Sensed Images*: An Introduction. John Willey and Sons Ltd., West Sussex, England.

MATLAB version 7.10.0. Natick, Massachusetts: The MathWorks Inc., 2010.

Maul, G. A. and Gordon, H. R. (1975). On the use of LANDSAT 1 in optical oceanography. *Rem. Sens. of Env.*, 4, 95 – 128.

Mayer, H., (1999). Automatic object extraction from aerial imagery-a survey focusing on buildings, *Computer Vision and Image Understanding*, 74 (2), 138–149.

Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines". *IEEE Transactions on Geoscience and Remote Sensing*, 42, 1778-1790.

Mitra, P., Shankar, B.U. (2004). Segmentation of multi-spectral remote sensing images using active support vector machines. *Pattern Recognition Letters* 25, 1067–1074.

Moussa, A. and El-Sheimy N. (2002). Man-Made Objects Classification from Satellite/Aerial Imagery Using Neural Networks. .http://www.isprs.org/proceedings/XXXVIII/part1/10/10_02_Paper_72.pdf.(retrieved 10/21/2015).

Mubea, K. and Menz, G. (2012). Monitoring Land-Use Change in Nakuru (Kenya) Using Multi-Sensor Satellite Data. *Advances in Remote Sensing*, 1, 74-84. doi:10.4236/ars.2012.13008 (http://www.SciRP.org/journal/ars)

Muller Klaus-Robert, Mika Sebastian, Rasch Gunnar, Koji Tsuda, and Schokopf Bern-hard (2001). An Introduction to Kernel-Based Learning Algorithms. *IEEE Transactions on Neural Networks*. 12(2), 181-201.

Nardinocchi, Kourosh Khoshelhama, Emanuele Frontoni, Carla, Mancini, Adriano and Zingaretti Primo, (2010). Performance evaluation of automated approaches to building detection in multi-source aerial data, Ancona, Italy. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65,123–133.

National Population Commission (2007): 2006 Population Census of the Federal Republic of Nigeria Official Gazzette, 94, (24).

Ndehedehe Christopher E (2014). Support Vector Machine Based Kernel Types in Extraction of Urban Areas in Uyo Metropolis from Remote Sensing Multispectral Image, *http://www.sciencepub.net/researcher*.(retrieved 10/21/2015) *6(2):105-112]*

Neelamegam S., and Ramaraj, E. (2003). Classification algorithm in Data mining: An Overview: *International Journal of P2P Network Trends and Technology (IJPTT)* 4 (8).

Obeta, M.C. (2015). Industrial Water Supply in Nnewi Urban Area of Anambra State, South Eastern Nigeria. *Journal of Geography, Environment and Earth Science International* 2(1): 12-23.

Ogbuagu, C.N, Okoli, U.J, Oguoma V.M, and Ogbuagu E.N. (2010). Orphans and Vulnerable Children Affected by Sexual Violence and HIV/AIDS in two Local Government Areas in Anambra State, Southeastern Nigeria. *American-Eurasian Journal of Scientific Research* 5(1):05-11.

Ohlhof T., Gülch E, Müller H, Wiedemann C, Torre M. (2015). Semi-Automatic Extraction of Line and Area Features from Aerial and Satellite Images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*,34,(XXX). http://www.isprs.org/proceedings/XXXV/congress/comm3/papers/316.pdf.

Ojiako, J.C. and Igbokwe, J.I. (2009). Application of Remote Sensing and Multimedia Geographic Information System (GIS) in the Administration of Socio-Economic Activities in Nnewi Urban Area of Anambra State, Nigeria. *Environmental Research Journal*, 3 (2), 60-67.

Okpala-Okaka, C., (2011). The Making of Urban Towns: A Comparison of the U.S and Nigerian Experiences. *Journal of Environmental Management and Safety*, 2. (1)75 – 82.

Okwuashi Onuwa, Jack McConchie, Peter Nwilo, Mfon Isong, Aniekan Eyoh, Okey Nwanekezie, Etim Eyo & Aniekan Danny Ekpo. (2012). Predicting Future Land Use Change Using Support Vector Machine Based GIS Cellular Automata: A Case of Lagos, Nigeria. *Journal of Sustainable Development Published by Canadian Center of Science and Education, 5, (5);* 1913-9071.

Onwutalobi, C.A (2009). History of Otolo Nnewi. http://www.codewit.com/africa/1235-history-of-otolo-nnewi-by-anthony-claret-onwutalobi. (Retrieved 10/23/2014).

Orajiaka, S. O. (1975). 'Geology' in Ofomata G.E.K. (ed.) *Nigeria in Maps: Eastern States*: Ethiope Publishing House, Benin City.

Orji Uzor E.N., and Obasi, S. N (2012). Properties and Classification of Erosion Prone Soils of Ukpor, Nnewi South L.G.A Anambra State Nigeria. *Int'l Journal of Agric. and Rural Dev. SAAT FUTO.*

Otukei, J.R., and Blaschke, T. (2010). Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms.

International Journal of Applied Earth Observation and Geoinformation 12S (2010) S27–S31

Owamoyo Najeem, A. Alaba Fadele, and Abimbola Abudu (2013). Number Plate Recognition for Nigerian Vehicles. *Journal of Academic Research International, Natural and Applied Sciences,* 4 (3),2223-9944.

Pal, M., & Mather, P.M. (2006). Some issues in the classification of DAIS hyperspectral data. *International Journal of Remote Sensing*, 27, 2895-2916.

Pal, M., (2007). Ensemble Learning with Decision Tree for Remote Sensing Classification. In Proceedings of the World Academy of Science, Engineering and Technology 26:735 – 737.

Persello, Claudio (2010). Advanced Techniques for the Classification of Very High Resolution and Hyperspectral Remote Sensing Images. *(Doctoral Dissertation*, The International Doctorate School in Information and Communication Technologies DISI-University of Toronto).

Phil-Eze, P.O (2010). Variability of Soil Properties related to Vegetation Cover in Tropical Rainforest Landscape. *Journal of Geography and Regional Planning,* 3 (7), 177-184.

Piera, J., Quesada, R., Manuel-lazaro, A., Del Rio, J., ShariatPanahi, S. and Olivar, G., (2004). Wavelet denoising technique for high-resolution CTD data. Characterization of turbulent oceanic flow," Sensors. *Proceedings of the IEEE*. 3: 1468-1471.

Pinho, C. M. D., Silva, F. C., Fonseca, L. M. C. and Monteiro, A. M. V. (2008). Intra-urban land cover classification from high-resolution images using the C4.5 algorithms. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*,37(B7),695-699.

Platt John C., (1998). Sequential Minimal Optimization. A Fast Algorithm for Training Support Vector Machines. *A Technical Report MSR-TR-98-14, April 21, 1998 Microsoft Research.*

Platt,J., (1999).*Probabilisticoutputsfor supportvectormachinesandcomparisonto regularizedlikelihoodmethods*.InSmola,A.J.,Bartlett,P.,Schölkopf,B.,Schuurmans, D. (eds.), Advances in Large Margin Classifiers, MIT Press, Cambridge, 61–74.

Plaza A, Benediktsson J.A, Boardman J.W., Brazile J, Bruzzone L, Camps-Valls G., Chanussot J, Fauvel M, Gamba P, Gualtieri A, Marconcini M, Tilton J.C, Trianni G, (2009). Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment,* doi: 10.1016/j.rse.2007.07.028.

Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazile, J., Bruzzone, L., Camps-valls, Pontil Massimiliano and Alessandro Verri, (1998). Support Vector Machines for 3D Object Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*Processing" Remote Sensing of Environment, 20. (6) 113, S110–S122.

Plaza, A., Martinez, P., Perez, R. and Plaza, J. (2002). Spatial/spectral endmember extraction by multidimensional morphological operations. *IEEE Trans. Geosci. & Remote Sensing*, 40, 2025-2041.

Plaza, A., Martinez, P., Perez, R. and Plaza, J. (2004b). A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. *IEEE Trans. Geosci. & Remote Sensing*, 42, (3), 650-663.

Plaza, A., Martinez, P., Perez, R.M. and Plaza, J. (2004a). A new approach to mixed pixel classification of hyperspectral imagery based on extended morphological profile, *Pattern Recognition*, 37, (6),1097 – 1116.

Plaza, A., Martínez, P., Plaza, J. and Pérez, R. (2003). Spectral analysis of hyperspectral image data. Advances in Technique for Analysis of Remotely Sensed Data, *IEEE Workshop*, pp. 298 – 307.

Plaza, A., Plaza, J. and Cristo, A. (2008). Morphological feature extraction and spectral unmixing of hyperspectral images. *IEEE International Symposium on Signal Processing and Information Technology*, pp. 12-17.

Pontil Massimiliano and Alessandro Verri, (1998). Support Vector Machines for 3D Object Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* VOL.20. NO. 6 Processing" Remote Sensing of Environment 113 (1), S110–S122.http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.3.8540.

Pradhan, B., Sharif, A., & Abubakar, A. (2013). Monitoring and predicting landuse change in Tripoli metropolitan city using cellular automata models in geographic information system. A paper presented at a seminar organized at the University of Cairo, Egypt.

Puissant, A., J. Hirsch, and C. Weber, (2005). The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution image.

Quinlan, R., (1993). *Programs for Machine Learning*. Morgan Kaufman, San Mateo.

R Core Team (2013). A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ,http://www.Rproject. orgwww.orfeotoolbox.org/otb/monteverdi.html (Available online, accessed 10/21/2015).

Radha, S. and Rehna, A.S. (2014). Retrieval of Geographic Images Using Multi Support Vector Machine and Local Invariant Features. *International Journal of Innovative Research in Science, Engineering and Technology an ISO 3297*: 2007 Certified Organization Volume 3.

Rafel, C., Gonzalez, Richard, E. &Woods, (2002). *Digital Image Processing*.2ndedition, Prentice Hall, Inc.

Rao, T., and Rajinikanth, T.V. (2014). Supervised Classification of Remote Sensed Data Using Support Vector Machine. *Global Journal of Computer Science and Technology: C. Software & Data Engineering*, Global Journals Inc. (USA)14 (1). 1.0.

Remote Sensing and GIS Laboratory, (2016). Digitizing and updating of existing maps Department of Environment Management, Chukwuemeka Odumegwu Ojukwu University, Uli-Campus.

Ren, R. E., and Wang, H. W., (1998). Multivariate Data Analysis (Theory, Method and Example). National Defence Industry Press (in Chinese).

Richards, J.A., (1993). *Remote Sensing Digital Image Analysis. An Introduction*. Springer-Verlag, Berlin.

Richards, J.A., (1999). *Remote Sensing Digital Image Analysis,* Springer-Verlag, Berlin, 240.

Roessner, S., Segl, K., Heiden, U., & Kaufmann, H. (2001). Automated differentiation of urban surfaces based on airborne hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing,* 39, 1525-1532.

Sabino (2005). Fundamentals of remote sensing. *Canadian Center for Remote Sensing*, pp. 148 & 161.

Sabins, F.F., (1997). Remote Sensing: Principles and Interpretation. W.H. Freeman and Company, New York.

Sali, E., Wolfson, H., (1992). Texture classification in aerial photographs and satellite data. *International Journal of Remote Sensing*, 13 (18), 3395-3408.

Sanchez, S., Martin, G., Plaza, A. and Chang, C. (2010). GPU implementation of fully constrained linear spectral unmixing for remotely sensed hyperspectral data exploitation. *In Proceedings SPIE Satellite Data Compression, Communications, and Processing* VI, (7810), 78100G-1 – 78100G-11.

Schokopf Bernhard, and Smola, Alexander J. (2002). *Learning with Kernels-support vector machines, regularization, optimization and beyond*. The MIT Press, Cambridge.

Schölkopf,B.,Burges,C.J.C.,andSmola,A.J., (1999).*AdvancesinKernelMethods*, MIT Press, Cambridge.

Shaw G.A., Burke H.K. (2003). Spectral imaging for remote sensing. *Lincoln Laboratory Journal*, 14, (1),3 – 28.

Shi Chun Yi, Ning Chang Huang, and He Jian Qin (2000). *Theory of Artificial Intelligent,* TSinghua University Press (in Chinese)

Shi Dee and Yang Xiaojun, (2012). Support Vector Machines for Landscape Mapping from Remote Sensor Imagery. *Proceedings-AutoCarto 2012 - Columbus, Ohio, USA* - September 16-18, 2012.

Singh S., Merugu, S., And Jain, K (2015). Land Information Extraction with Boundary Preservation for High Resolution Satellite Image. *International Journal of Computer Applications 120(.7),* 39

Sitaram, B. and Manjunath B. S. (2006). Modeling and Detection of Geospatial Objects Using Texture Motifs" *IEEE Transactions on Geoscience and Remote Sensing,* Vol. 44, Issue: 12, Pp. 3706 – 3715. DOI: 10.1109/TGRS.2006.881741.

Smola, A.J., (1996). *Regression estimationwithsupport vector learning machines*. (Master's thesis). Technique University of Munich.

Smola,A.J.,Bartlett,P.,Schölkopf,B.,andSchuurmans,C., (1999).*AdvancesinLarge Margin Classifiers,* MIT Press, Cambridge.

Soludo C.C., (2006). Anambra 2030: Envisioning the African Dubai, Taiwan and Silicon Valley. *A keynote Address presented at the 2nd Anambra State Economic Summit at Parktonia Hotel, Awka* May 25- 26, 2006.

Song Qing, Hu Wenjie, Xie Wenfang (2002). Robust support vector machine with bullet hole image classification". *IEEE Trans SystCybern*, 32(4):440-448.

Steele, B., M. and Patterson, D. (2001). A Land Cover Mapping using Combination and Ensemble Classifiers. *Computing Science and Statistics*,33, 236 – 247.

Su, L., Chopping, M.J., Rango, A., Martonchik, J.V., and Peters, D.P.C., (2007). Support Vector Machines for recognition of semi-arid vegetation types using MISR multi-angle imagery. *Remote Sensing of Environment*, 107, (1-2), 299-311.

Sumengen, S. Bhagavathy, and Manjunath,B. S. (2004). Graph partitioning active contours for knowledge-based geospatial segmentation in *Proceedings of the IEEE CVPR Workshop on Perceptual Organization in Computer Vision*, pp. 54–54.

Suresh Singh, Merugu Suresh, Jain K., (2015). Land Information Extraction with Boundary Preservation for High Resolution Satellite Image. *International Journal of Computer Applications,* 120, (7), 0975 – 8887

Suykens,J.A.K.,andVandewalle,J., (1999).Leastsquaressupportvectormachine classifiers.*Neural Processing Letters*, 9(3), 293–300.

Suykens, J.A.K. Brabanter, J., Lukas, De L Vandewalle, J. (2001). Weighted Least Squares Support Vector Machines: *Robustness and Sparse Approximation. Neurocomputing* 48 (2002) 85–105.

Taha, H. A., (1997). *Operations research: an introduction.* Prentice Hall, New Jersey.

Takuya Inoue, and Shigeo Abe (2001). Fuzzy support vector machines for pattern classification. *In Proceedings of International Joint Conference on Neural Networks*, 2, 1449-1454.

Transparency International: (http://www.transparency.org/).(Available online, accessed 10/21/2015).

Tsai F. and Philpot, W. D. (2002). A derivative-aided hyperspectral image analysis system for land-cover classification. *IEEE Trans. Geosc. Remote. Sensing*, 40, 416–425.

Tsai F., Chou M.J., Wang H.H. (2005) -Texture Analysis of High Resolution Satellite Imagery for Mapping Invasive Plants. *Geoscience and Remote Sensing Symposium, 2005. IGARSS '05 Proceedings. 2005 IEEE International,* 4, 3024-3027. DOI: 10.1109/IGARSS.2005.1525707

UCI, repository of bioinformatics Databases, Website: http://www.ics.uci.edu/~mlearn/MLRepository.html.

Ulasi, E., Ayadikwor, C., Ejiofor, P., Chidulue, C., and Achuwa, N. (2017). Personal Communications.

UN-HABITAT. (2009). *Structure Plan for Anambra State. Nairobi, Kenya*: United Nations Human Settlements Programme Publishers.

Ünsalan, C., Boyer, K.L., (2005). A system to detect houses and residential street networks in multispectral satellite images. *Computer Vision and Image Understanding*, 98 (3), 423–461.

Ustuner Mustafa, Fusun BalikSanli and Barnali Dixon., (2015). Application of Support Vector Machines for Landuse Classification Using High-Resolution Rapid Eye Images: A Sensitivity Analysis. *European Journal of Remote Sensing*, 48, 403-422.

Van der Linden, S., & Hostert, P. (2009). The influence of urban structures on impervious surface maps from airborne hyperspectral data. *Remote Sensing of Environment*, 113, 2298-2305.

Vapnik Vladimir N. (1995). *The Nature of Statistical Learning Theory*. Springer, Berlin Heidelberg New York.

Vapnik, V. (1999). *The nature of statistical learning theory*. (2nd ed). New York: Springer-Verlag.

Vapnik, V. and A. Lerner, (1963). Pattern recognition using generalized portrait method. *Automation and Remote Control*. Vol.24, 774-780.

Vapnik, V., (1998). Statistical Learning Theory. New York: Wiley.

Vapnik, V.N., Chervonenkis, A.Y., (1971). On the uniform convergence of the relative frequencies of events to their probabilities. *Theory of Probability and its Applications* 17, 264–280.

Varshney, P. K., and Arora, M. K. (2004). *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*. Springer-Verlag Berlin Heidelberg,eBook ISBN978-3-662-05605-9, XV, 323,

Verbeke, L.P.C., Vabcoillie, F.M.B., (2004). Re-using back propagating artificial neural network for land cover classification in tropical savannahs. *International Journal of Remote Sensing* 35, 2747–2771.

Vesanto, J., Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks,* 11(3) 586–600.

Wacker, A. G., and Landgrebe, D. (1972). Minimum distance classification in remote sensing. LARS *Technical Reports. Paper 25.* Online publication. http://docs.lib.purdue.edu/larstech/25 Cited 11 July 2015.

Weidner, U., and Förstner, W., (1995). Towards automatic building extraction from high resolution digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing,* 50(4), 38–49.

Wiedemann, C., Heipke, C., Mayer, H., and Jamet O., (1998). *Empirical evaluation of automatically extracted road axes, in Empirical Evaluation Methods in Computer Vision,* J. B. Kevin and P. J. Phillips, Eds. Silver Spring, MD: IEEE Computer Society Press, 172-187.

Wolter, J.A. (1975). *The Emerging Discipline of Cartography, Minneapolis:(* Doctoral dissertation. The University of Minnesota).

Xia Hong and Bao Hu Qing (2005): “Feature selection using fuzzy support vector machines. *Fuzzy Optimal Decision Making*, 2006, 5(2):187-192.

Xian Si-dong (2010). A new fuzzy comprehensive evaluation model based on the support vector machine. *Fuzzy Information and Engineering*, 2(1), 75-86.

Xie, Y., Sha, Z.and Yu, M., (2008). Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology*, 1, 9–23.

Xie, Chenglin (2006). *Support Vector Machines for Land Use Change Modelling.* (Master's Thesis) Department of Geomatics Engineering Calgary, Alberta.

Xin, H., Dziadek, S., Bundle, D. R. & Cutler, J. E. (2008). Synthetic glycopeptide vaccines combining b-mannan and peptide epitopes induce protection against candidiasis. *Proc Natl Acad Sci U.S.A*, 10(5), 13526–13531.

Xiufeng Jiang, Zhang Yi, Jian Cheng Lv. (2006). Fuzzy SVM with a new fuzzy membership function. *Neural Computer Applications*, 15: 268-276.

Yang, Xiaojun (2011), Parameterizing Support Vector Machines for Land Cover Classification, *American Society Photogrammetric Engineering & Remote Sensing*, 77 (1), 27–37.

Yao Yukai, Yang Liu, Yongqing Yu, Hong Xu, WeimingLv, Li Zhao, Chen Xiaoyun, (2013). "K-SVM: An Effective SVM Algorithm Based on K-means Clustering. *Journal of Computers*, 8(10), 2632-2639.

Yu Hwanjo, Yang Jiong and Han Jiawei (2003). Classifying Large Data Sets Using SVMs with Hierarchical Cluster. SIGKDD 2003. Washington, DC, USA.

Zhang Jixian, Lin Xiangguo and Ning Xiaogang, (2013). SVM-Based Classification of Segmented Airborne LiDAR Point Clouds in Urban Areas. *Journal of Remote Sensing*, 5, 3749-3775.

Zhang Q, and Wang J. (2001). Texture analysis for urban spatial pattern study using SPOT imagery. *IEEE T Geosci Remote.;* 28(4):513-519.

Zhang Xuegong (1999). Using class-center vectors to build support vector machines. *In Neural Networks for Signal Processing IX, New York: The Institute of Electrical and Electronics Engineering*, Inc. 1999, 3-11.

Zhang Yong, Chi Zhong-xian, Liu Xiao-Dan, and Wang Xiang-Hai (2007). A novel fuzzy compensation multi-class support vector machines". *Applied Intelligence*, 27(1), 21-28.

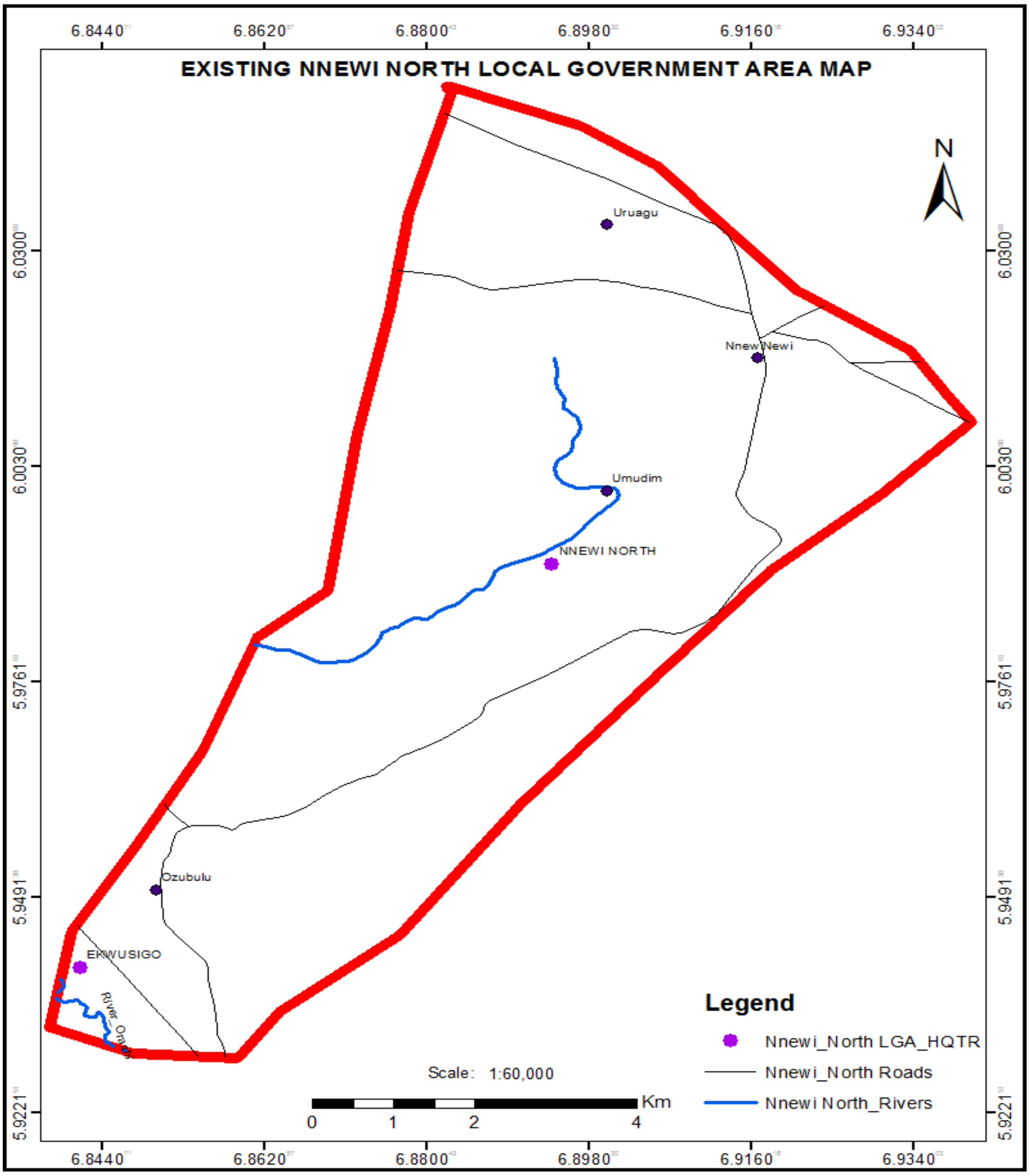
Zhang, L. and Huang, X. (2010). Object-oriented subspace analysis for airborne hyperspectral remote sensing imagery. *Neuro-computing*, 73, 927-936.

Zhang, Y.S., (1996). A hierarchical neural network approach to three-dimensional object recognition. *International Archives of Photogrammetry and Remote Sensing, XXXI* (B3): 1010-1017.

Zhu C. and Yang, X. (1998). Study of remote sensing image texture analysis and classification using wavelet. *International Journal of Remote Sensing*, 19 (16), 3197–3203.

APPENDIXA1

OLD SHAPE FILE OF NNEWI NORTH LGA

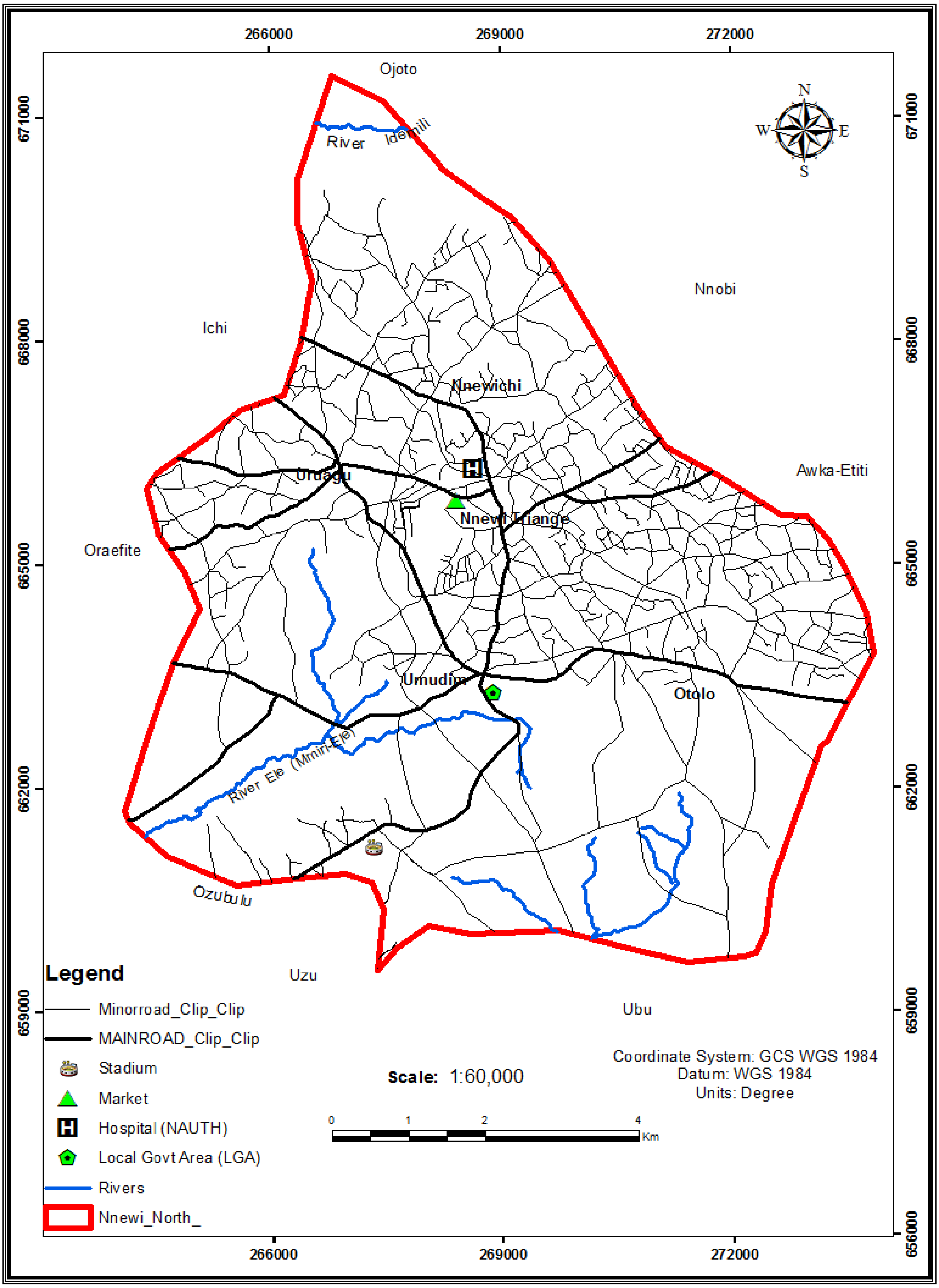


APPENDIX A1: Showing old Nnewi Local Government Area map comprising of some part of Ekwusigo LGA. This Shows that the Nnewi-North LGA map is not updated.

Remember more States and Local government areas were created in Nigeria in 1967, 1976,1987, 1991and 1996 respectively and subsequently some of our administrative boundaries are not updated and one of them is this map of old Nnewi Local Government Area.

APPENDIX A2

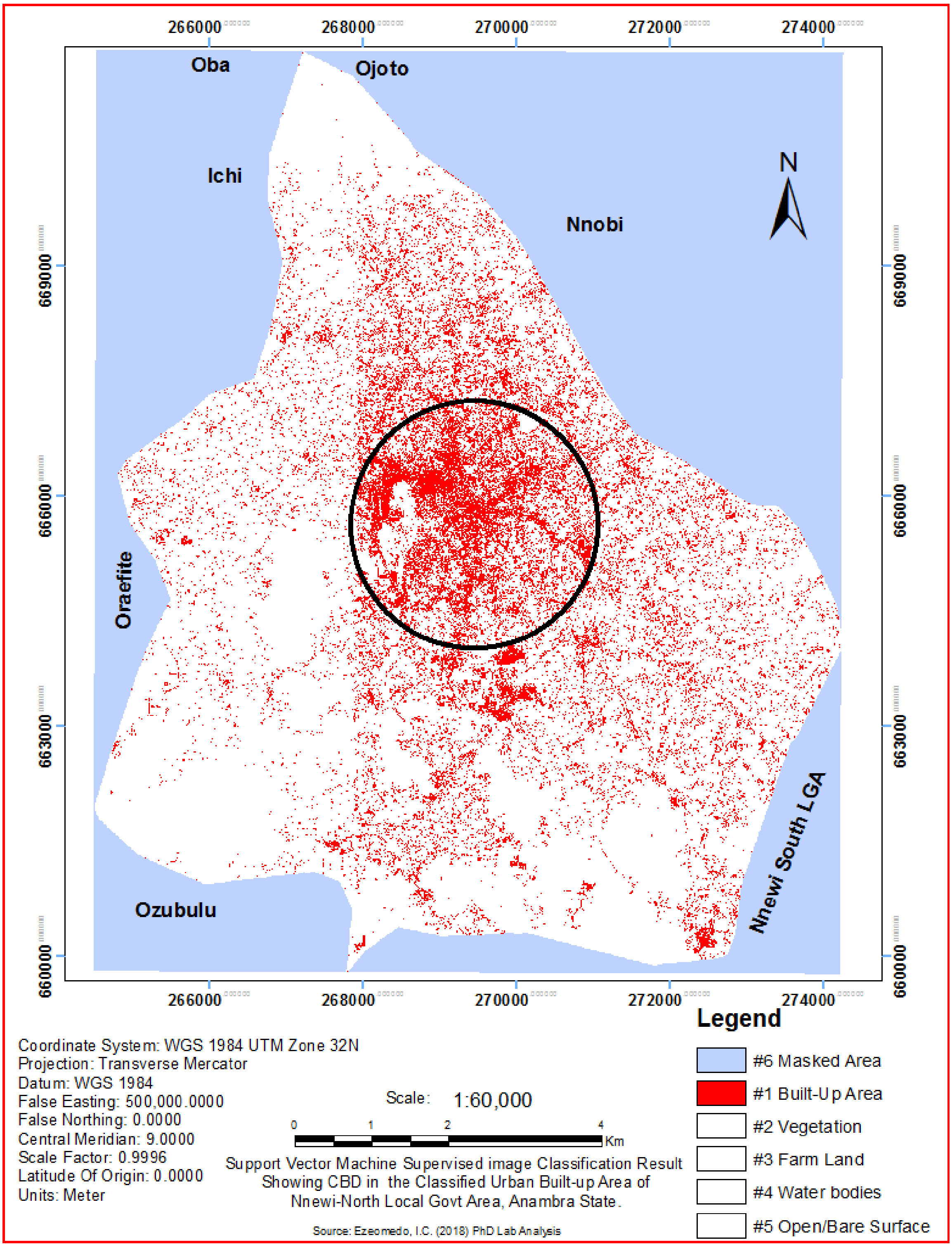
NEW SHAPE FILE OF NNEWI NORTH LGA



APPENDIX A2: The current precise boundary of Nnewi-North was used to clip the satellite imagery covering Nnewi-North LGA. This shape was used in image classification and map production.

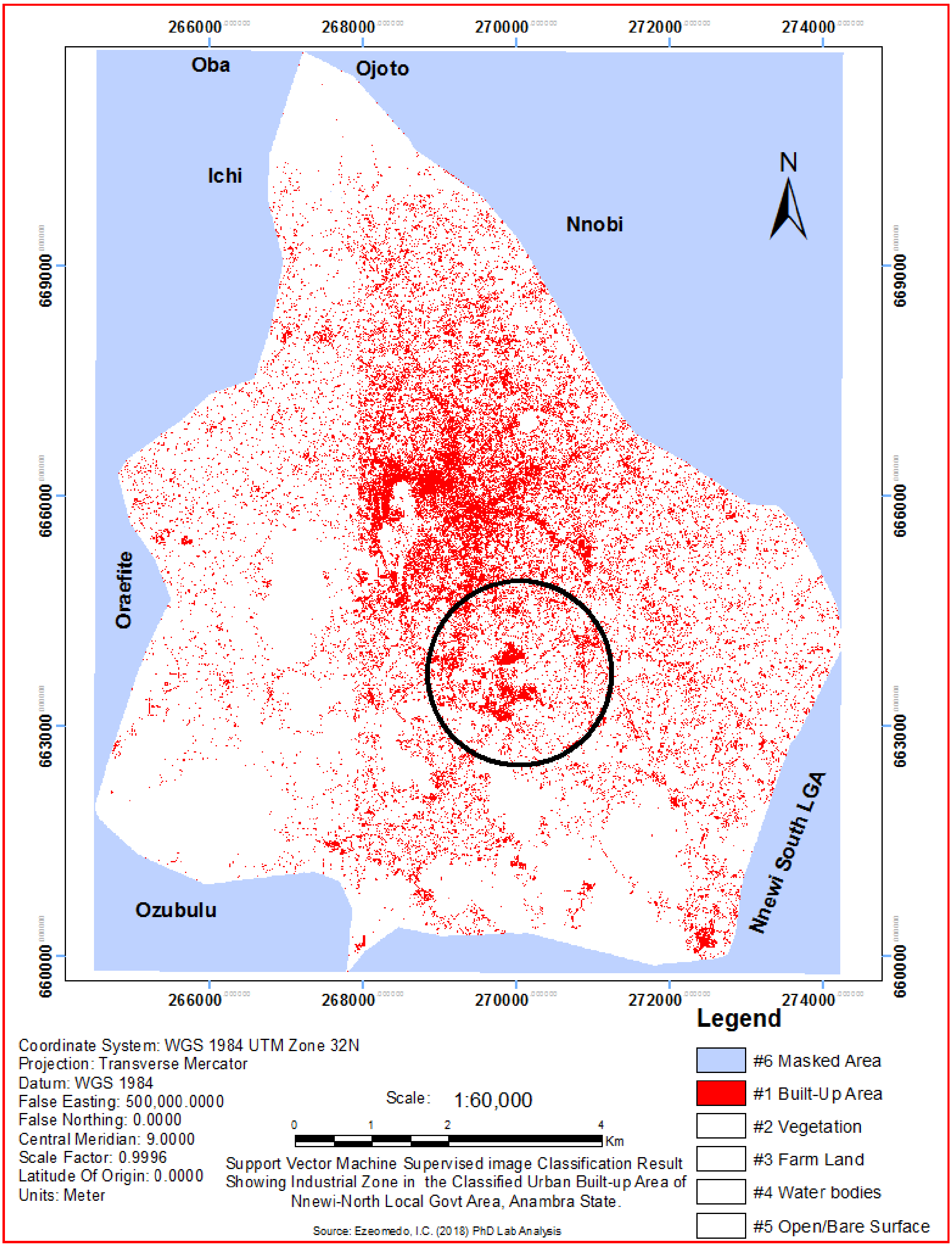
APPENDIX B1:

SPECIFIC CLASSIFICATION RESULT



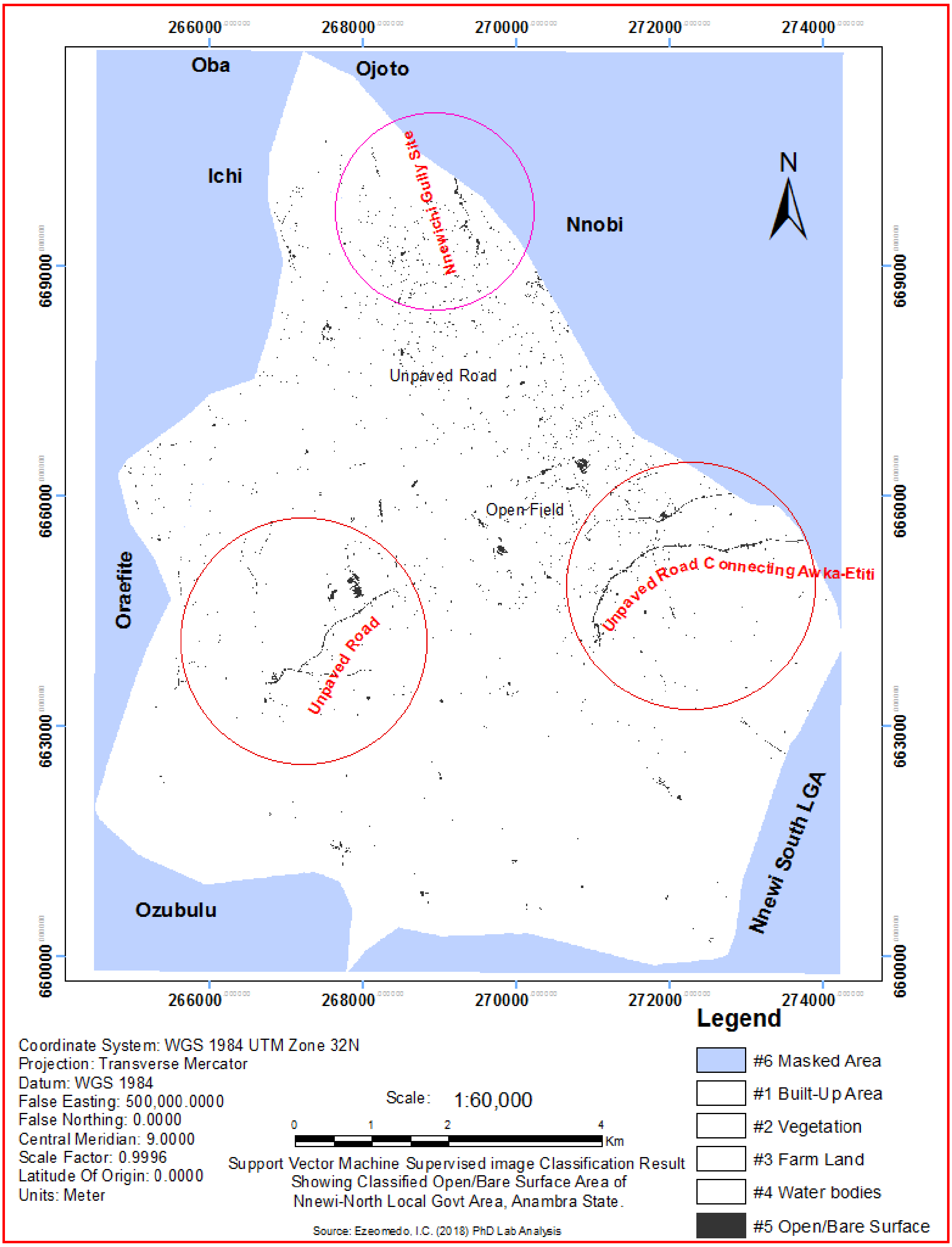
APPENDIX B1: Classified Built-Up Area indicating Nnewi-North LGA Central Business Districts (CBD) the popular 'Nkwo Nnewi' encircled inblack ink. The result shows a compacted Settlement pattern.

APPENDIX B2



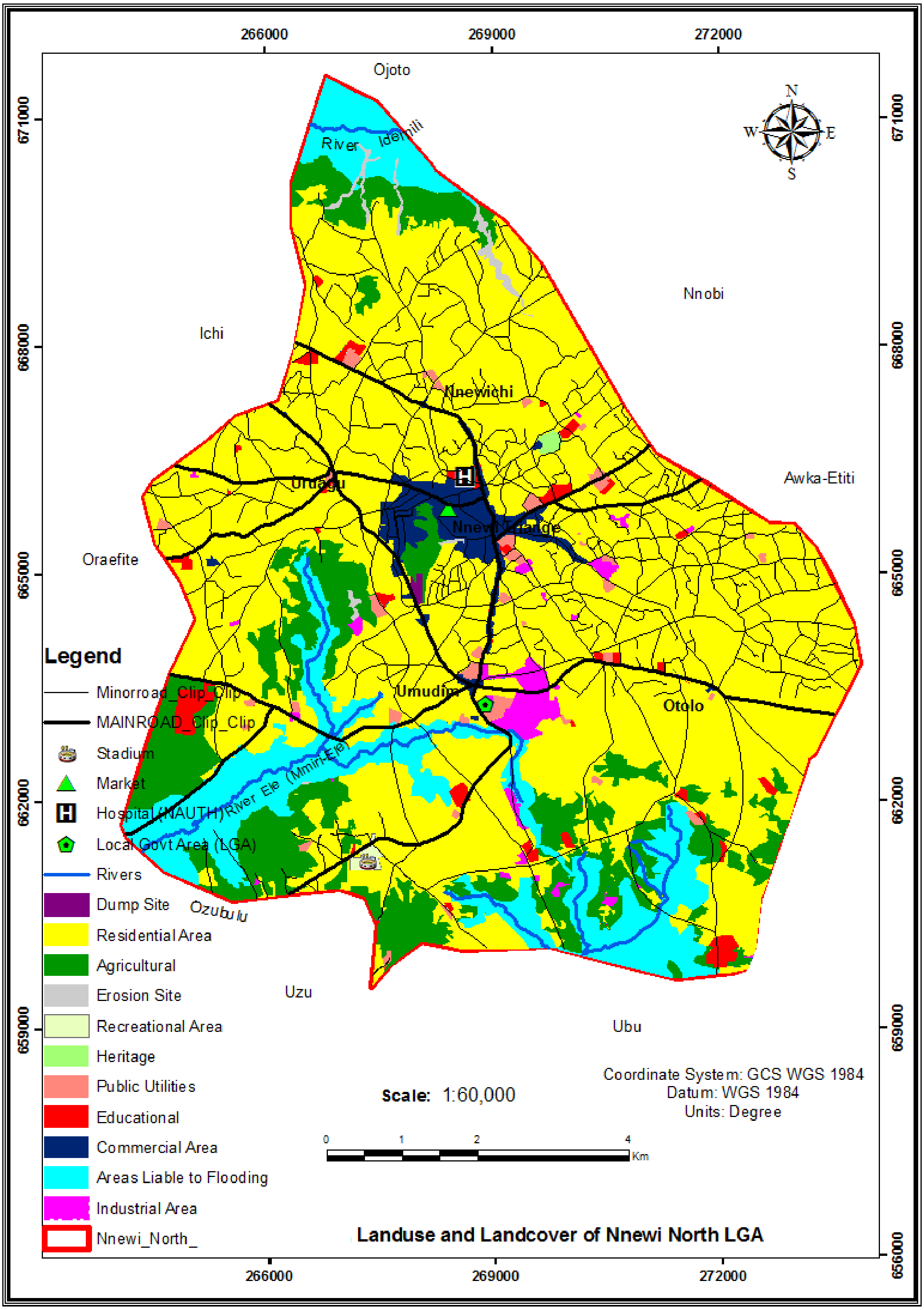
APPENDIX B2: Classified Built-Up Area indicating Nnewi-North LGA Industrial Zone mainly at Umudim Industrial Estate,encircled in black ink. The presence of an industrial estate and a big market "Nkwo Nnewi' is believed to have brought the development to Nnewi and therefore attracted huge population to the city.

APPENDIX B3



APPENDIX B3: Classified Open/bare Surface. The highlight in red circle indicates how well unpaved roads, bare surfaces and erosion site were classified using ENVI.

APPENDIX B 4



APPENDIX B4: Physical vector map of Nnewi North LGA showing the Landuse/landcover