

# Generating Land Use and Land Cover Maps with Open Data Using Open-Source Tools: Forecariah, Guinea Case Study

Submitted to the Graduate School of Natural and Applied Sciences in partial  
fulfillment of the requirements for the degree of

Master of Science

in Geomatics Engineering

by

Youssouf Camara

ORCID 0000-0001-7262-6713

October 2021

This is to certify that we have read the thesis **Generating Land Use and Land Cover Maps with Open Data Using Open-Source Tools: Forecariah, Guinea Case Study** submitted by **Yousseuf Camara**, and it has been judged to be successful, in scope and in quality, at the defense exam and accepted by our jury as a MASTER'S THESIS.

**APPROVED BY:**

**Advisor:**                      **Assist. Prof. Dr. Mehmet Güven Koçak**                      .....

İzmir Kâtip Çelebi University

**Committee Members:**

**Assist. Prof. Dr. Serkan Karakış**                      .....

İzmir Kâtip Çelebi University

**Assist. Prof. Dr. Kutun Koruyan**                      .....

Dokuz Eylül University

**Date of Defense: August 12, 2021**

# Declaration of Authorship

I, **Youssef Camara**, declare that this thesis titled **Generating Land Use and Land Cover Maps with Open Data Using Open-Source Tools: Forecariah, Guinea Case Study** and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for the Master's degree at this university.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this university or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. This thesis is entirely my own work, with the exception of such quotations.
- I have acknowledged all major sources of assistance.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signature:

---

Date: 26.10.2021

---

# Generating Land Use and Land Cover Maps with Open Data Using Open-Source Tools: Forecariah, Guinea Case Study

## Abstract

Due to anthropological actions, land usage changes within time. Those changes impact both land management and environment safe keeping regulations which therefore adverse living conditions. Tracking those changes have become the backbone for developing huge engineering projects and capital decision making processes. Especially in developing countries, where financial resources are very limited. Through remote sensors and open access programs, LULC features motoring can be performed with some ease for such countries. Sentinel 2 MSI with its 10m ground sampling and 5 days revisit frequency can be used to perform such study. We used Sentinel 2 image of Forecariah, a region with 585 sqkm and a population around 240000, located in a close proximity to Conakry the capital city of Guinea recognized to be the most mass migration center of the country. Globeland30 was used to classify the area LULC features such as water, wetland, agriculture, vegetation, bare-land, built-up. K-means was performed as unsupervised method and RF, SVM, DL4J were used as supervised method for finding the clustering algorithm that best fit our purpose and can be used free of charge. Iteration number was set to 30 and class limit to 6 for the unsupervised method. For the supervised case, Google Earth image of data were used to set train and test polygons, assessed with 10-fold cross validation. As a result, thematic maps were obtained. K-means was run under SNAP, the sentinel data processing software, RF and SVM under EnMAP-Box a python plugin for QGIS and DL4J under WEKA which provides access to both SQL database and DL through DL4J. As

overall accuracies, we obtained 78%, 83%, 86% and 91% respectively for K-means, RF, SVM and DL4J. DL4J gave best result as expected because relevant to DL compared to other 3 classifiers. Therefore, we can advise the usage of DL algorithms in engineering and environment protecting projects as well as in the building of a decision making systems. Especially, when it comes to monitoring LULC features change that can impact the life of future generations and mostly in low-income regions.

**Keywords:** Deep Learning (DL), Support vector Machine (SVM), Random Forest (RF), K-mean, QGIS, Weka, SNAP, LULC, decision making systems

# Açık Görüntü Verilerinden Açık Kaynak Araçlar Kullanılarak Arazi Kullanım ve Arazi Örtüsü Haritalarının Üretilmesi: Forecariah, Gine Örneği

## ÖZ

Antropolojik eylemler nedeniyle arazi kullanımını zaman içinde değiştirmektedir. Bu değişiklikler hem arazi yönetimini hem de çevre koruma düzenlemelerini etkiler ve dolayısıyla yaşam koşullarını olumsuz yönde değiştirir. Bu değişiklikleri izlemek, devasa mühendislik projeleri ve sermaye karar verme süreçleri geliştirmenin bel kemiği haline gelmiştir. Özellikle finansal kaynakların çok sınırlı olduğu gelişmekte olan ülkelerde. Uzaktan algılayıcı sensörler ve açık erişim programları aracılığıyla, AKAÖ özellikleri bu tür ülkeler için bir miktar kolaylıkla gerçekleştirilebilir. Sentinel 2 MSI, 10m yer örnekleme aralığı ve 5 günlük zamansal çözünürlük ile bu tür bir çalışmayı gerçekleştirmek için kullanılabilir. Ülkenin göç merkezi olarak kabul edilen Gine'nin başkenti Conakry'ye yakın bir yerde bulunan 585 km<sup>2</sup>'lik ve 240000 civarında nüfusa sahip bir bölge olan Forecariah'ın Sentinel 2 görüntüsü kullanılmıştır. Globeland30, su, sulak alan, tarım, bitki örtüsü, çıplak arazi, yapılaşma gibi alan AKAÖ özelliklerini sınıflandırmak için kullanılmıştır. Amacımıza en uygun ve ücretsiz olarak kullanılacak kümeleme algoritmasını bulmak için denetimsiz yöntem olarak K-means, denetimli yöntem olarak RF, SVM, DL4J kullanıldı. Denetimsiz yöntem için yineleme sayısı 30 ve sınıf sınırı 6 olarak ayarlandı. Çalışmada 10 kat çapraz doğrulama ile yapılmış, eğitim ve test poligonlarını seçmek için Google Earth görüntüsü kullanılmıştır. Bunun sonucunda tematik haritalar elde edilmiştir. K-means, Sentinel veri işleme yazılımı SNAP altında çalıştırılmıştır, RF ve SVM için

EnMAP-Box ve DL için WEKA kullanılmıştır. Genel doğruluk olarak K-means, RF, SVM ve DL4J için sırasıyla %78, %83, %86 ve %91 elde edilmiştir. Diğer 3 sınıflandırıcıya kıyasla DL ile ilgili olduğu için beklendiği gibi en iyi sonucu vermiştir. Bu nedenle, mühendislik ve çevre koruma projelerinde ve güçlü karar verme sistemlerinin oluşturulmasında DL algoritmalarının kullanılmasını önerebiliriz. Özellikle, gelecek nesillerin yaşamını etkileyebilecek ve çoğunlukla düşük gelirli bölgelerde LULC özelliklerinin değişmesinin izlenmesi söz konusu olduğunda açık kaynak veri ve yazılımlardan yararlanılması tavsiye edilmektedir.

**Anahtar Kelimeler:** Derin Öğrenme (DL), Destek Karar Makinesi (SVM), Rastgele Orman (RF), K-means, QGIS, Weka, SNAP, AKAÖ, karar Destek Sistemleri

*To those who have been taken away from our affections, those who left our world earlier while we are away from them, I here think about: My beloved younger sister Hadja Kadiatou CAMARA*

*GOD bless her soul and keep her in paradise.*

*And to those who have suffered from my absence: My family and friends.*

***To my family***



# Acknowledgment

First and foremost, I would like to thank my advisor, Dr. Mehmet Guven KOÇAK for his invaluable support and technical discussions that I have always appreciated. Beside him, all the professors of Geomatics Engineering Department and the Faculty of Applied and Natural Sciences at İzmir Katip Çelebi University for their guidance, sincere cooperation, positive attitude, and understanding during my BSc and MSc education. To these elegant people of İzmir Katip Çelebi University, my recognition and everlasting remembrance. Beside them, I would like to dedicate this work to many people for having supported and facilitating my stay in Turkey and my studies in İzmir Katip Çelebi University. Within these people, in the first rank, comes my family father Ibrahima, mother Halimatou Doumbouya, brothers and sisters for their valuable support and encouragement during my whole student life and this thesis preparation. Then, comes my respectable colleague Metin Balcı, friends Moussa Kalil Fofana, Kadiatou Diallo, Tidiane Sacko, Mariama Dalanda Diallo, Mohamed Rachid Fofana, Loic Djana, Mamadou Alimou Diallo, Daouda Konate, Hassata Diallo, Djibrilla Harouna Garba, Patson Vivelea. I also would like to thank once again to Dr. Mehmet Guven KOÇAK my advisor, who wished and expected me to work harder and never to give up, so that I could be one of his first full time mastering MSc student to be graduated under his coaching at İzmir Katip Çelebi University. He has carefully read and corrected the full volume. His many suggestions and improvements, critical remarks and proposals are gratefully acknowledged. To him, my endless recognitions and respect. Finally, I would like to extend my thanks and appreciation to the Director of the Graduate School of Natural and Applied Sciences of İzmir Katip Çelebi University Prof. Dr. Mehmet ÇEVIK, for all his efforts to wellbeing of all international students at the University and Izmir. This thesis was supported by Izmir Katip Celebi University, Coordination Office of Scientific Research Projects.

# Table of Contents

Declaration of Authorship .....	iii
Abstract .....	iii
Öz .....	vi
Acknowledgment .....	ix
List of Figures .....	xii
List of Tables .....	xiii
List of Abbreviations .....	xiv
<b>1 Introduction .....</b>	<b>1</b>
1.1 Problem Statement .....	4
1.2 Research Objective .....	5
1.3 Thesis Outlines .....	5
<b>2 Land Use Land Cover Classification Methods .....</b>	<b>6</b>
2.1 Land Use Land Cover Classification Systems .....	6
2.1.1 Unsupervised Classification .....	7
2.1.1.1 K-Means Clusters Analysis .....	7
2.1.2 Supervised Classification .....	8
2.1.2.1 Random Forest .....	11
2.1.2.2 Support Vector Machine .....	13
2.1.2.3 Deep Learning .....	14
<b>3 Application of LULC Classification in Forecariuah .....</b>	<b>17</b>
3.1 Study Area and Data .....	21

3.1.1 Study Area.....	4
3.1.2 Data .....	4
3.1.2.1 K-Means Clusters Analysis .....	14
3.1.2.2 Random Forest .....	14
3.1.2.1 Support Vector Machine .....	14
3.1.2.1 Deep Learning For Java .....	14
3.2 Pros and Cons .....	15
3.2.1 Results .....	15
3.2.2 Discussions.....	16
<b>4 Conclusion and Recommendations .....</b>	<b>17</b>
4.1 Conclusion .....	25
4.2 Recommendations .....	25
<b>References .....</b>	<b>27</b>
<b>Curriculum Vitae .....</b>	<b>38</b>

# List of Figures

Figure 2.1 Decision Tree .....	5
Figure 2.2 Support vector Machine .....	7
Figure 2.3 Deep Learning Neural Network Processing Steps .....	10
Figure 3.1 Map of Guinea .....	11
Figure 3.2 Map of Forecariah region .....	12
Figure 3.3 Sentienl 2A image .....	15
Figure 3.4 Subsetting process of raster data .....	16
Figure 3.5 Subsetted Raster data .....	18
Figure 3.6 Radiometric Calibration.....	20
Figure 3.7 Geometric Calibration .....	21
Figure 3.8 Region Of Interest .....	22
Figure 3.9 K-Means processing .....	23
Figure 3.10 Number of clusters and iteration setting .....	24
Figure 3.11 Default clusters and iteration values .....	25
Figure 3.12 K-Means classification map .....	27
Figure 3.13 RF classification method settings .....	28
Figure 3.14 RF classification map .....	29
Figure 3.15 SVM classification method settings .....	29
Figure 3.16 SVM classification map .....	29
Figure 3.17 Weka DL environment .....	29
Figure 3.18 DL4J classification map .....	29

# List of Tables

Table 3.1	K-Means Clusters Analysis results .....	6
Table 3.2	RF cross-validation results .....	9
Table 3.3	SVM cross-validation results .....	11
Table 3.4	DL4J cross-validation results .....	14
Table 3.5	Classifications results .....	15
Table 3.6	Class-wise Accuracies .....	16

# List of Abbreviations

AI	Artificial Intelligence
AM	Aerial Mapping
ARFF	Attribute Relational File Format
DL	Deep Learning
DT	Decision trees
EU	European Union
GIS	Geospatial Information Sciences
GNSS	Global Navigation Satellite Systems
GPS	Global Positioning System
IKCU	Izmir Katip Celebi University
K-Means	K-Means Clustering Analysis
LA	Land Administration
LM	Land Management
LS	Land Surveying
LULC	Land Use Land Cover
ML	Machine Learning
NASA	National Aero-Space Administration
NCES	National Center for Space Studies
NIR	Near-Infrared
NN	Neural network
QGIS	Quantum Geographic Information Sciences
RF	Random Forest
RS	Remote Sensing

SNAP	The Sentinel Application Platform
SVM	Support Vector Machine
SWH	Spot World Heritage
TS	Terrestrial Surveying
USGS	United States Geological Survey
WEKA	Waikato Environment Knowledge Analysis

# Chapter 1

## Introduction

Land Use Land Cover (LULC) classification is a process of clustering features. It has become the backbone of several aspects of land management (LM) that directly impacts our social and economic life. LULC classification can be applied e.g., for monitoring the spatial and temporal expansion of a city, studying climate change, tracking land degradation, controlling marine-related activities. It must strongly guide our daily activities and initiate the build-up of a solid decision-making system.

Although it is difficult to define exactly what constitute the Land Use, it may refer to human activity that is directly related to the land. Land Cover then would describe the natural and artificial features that exist onto the land surface [1]. LULC change tracking allows one to keep permanently all the details about the usage on land. Both positive and negative impacts of those usages are to be evaluated and different alternative solutions should be generated for better decision-making. It should control all the changes that have been taking place within a location and its proximity. Objective decisions can be made in planning studies for future generations by using its quantitative output. LULC change tracking has been gaining importance especially due to permanent and rapid changes in land surface features. Monitoring those features has become a must for governments willing to gather data and increase their revenues accordingly. On the other hand, land administration (LA) also needs actual LULC data to carry out rural land reforms, improve urban planning, provide infrastructure development, and support environmental monitoring [2]. Thus, information extracted from LULC will establish a strong base for a realistic and sustainable LA.

Classifying LULC features within time has become a very useful topic for decision-makers and it can be carried out at different scales. It is an important and indispensable topic in Geomatics engineering,



especially for photogrammetry and remote sensing (RS) subfields. Classification is the categorization of features and their labeling based on the reflectance or pixel values of objects in a digital image.

For terrestrial geodetic surveying, surveyors go to the fields with their instruments and tools, get into physical contact with objects. They perform data collection and measurements based on the purposes assigned to them. For LULC determination, all land cover features are recorded into total stations, GNSS receivers together with a sketch. Features may cover characteristic points for drawing contour lines, parcel and building corners, intersection points, etc., and their point-wise coordinate values are stored for later use. Real-time intervention may be necessary for transportation/route planning surveys since points obtained from a previously generated plan must be stacked out in the field immediately. Terrestrial methods are mostly used for their high precision and site practical feasibilities. They can also be incorporated into machine control working systems like road construction.

In former times, terrestrial methods were mainly used to build a geodetic infrastructure throughout a country. The idea behind building networks across the country was to create topographic maps which in turn have been base for larger-scale projects like cadaster, land readjustment, and consolidation. The chain of technical developments from the early astronomical surveyors to the present satellite geodesists reflects man's desire to be able to master time and space and to use science to foster society. The surveyor's role in society has remained unchanged from the earliest days. That is to determine land boundaries, provide maps of his environment, and control the construction of public and private works [3]. Even though terrestrial surveys provide precise high accurate output they are slow due to their very nature when it comes to covering larger areas such as countries and regions compared to other more recent modern methods. They need not only outdoor investment in time and cost but also in some office work for post-processing where in-field-collected data are analyzed.

Nowadays, terrestrial observations have been still carried out for specific and small projects covering a limited area. However, depending on the scale remotely sensed (RS) imagery with varying resolution will be rather preferred for larger projects. For handling complex urban and environmental problems, multidisciplinary datasets provided by new technologies is required. These technologies despite data they provide, make available tools and computer programs (SNAP, EnMAP-Box,...) to analyze provided data. They help to monitor the dynamic, natural and physical processes that affect the earth and adverse living conditions, for better decision making purpose[4]. RS methods use satellite and drone images to extract LULC features. This helps in the full representation process of

the earth's topography. Emphasis is made on modern techniques and tools that are all materialized in algorithms and computer programs to obtain an implementation-ready output. Different from terrestrial geodetic methods RS imagery can be utilized by a wide group of disciplines like urban studies, agriculture, geosciences, forestry, geology, geography, environmental studies, etc. Professionals from these fields also use GIS to integrate classification maps obtained by RS imagery and tools. They exploit several important features of GIS to analyze and carry out necessary operations on the classification map. Although computer algorithms and models have advanced to a mature level their users still need to intervene for making reasonable deductions and taking further actions [5]. Thus, information extraction from RS imagery in terms of LULC by using modern tools gains an ever-increasing importance for synthesizing it with other layers of information and even for knowledge discovery. LULC mapping is therefore an essential stage before getting into contact with other types of data/information under a decision-making system [6].

Governments and policy makers can benefit from the power of decision-making systems and the products of new technologies to monitor changes in LULC. They can use above mentioned tools to develop projects, unite planning efforts and implement adopted plans successfully. Besides, tools provide managers a holistic view for making and executing LA decisions. Sustainable LA requires observing legal and illegal activities on land [7]. A careful monitoring and examining can be performed by extracting LULC from a sequence of RS images. Thus, approved legal activities like construction project planning, land readjustment and consolidation and illegal activities like deforestation, wildfire, land occupation can be kept under control. Rapid response to the observed changes can only be evoked by adopting efficient tools and technology. Tracking those changes is performed in several ways depending on the development level of countries.

RS depends entirely on images collected by satellites, airplanes, and increasingly drones. End users get the choice to select data to be processed based on the type of application which in turn specifies the spatial and spectral resolution of the imagery. Conducting any RS related project needs some monetary investment and image purchasing costs comprise the majority of the project's total expenses. Especially low-income countries experience financial difficulty in obtaining RS imagery. Concerning image costs, users have entered in January 2008 into a new era in which USGS (United States Geological Survey) and NASA (national Aero-Spatial Agency) adopted free and open data policy [8]. Landsat's huge multispectral data have been made accessible for everyone with its 30 m spatial and 16-day temporal resolution. European Commission has also declared free and open data

policy within its Copernicus program [9]. Multispectral imagery from Sentinel-2's twin satellites can be obtained freely with varying spatial resolution between 10 to 60 m and high revisit frequency of 5 days [10]. Lately, French National Centre for Space Studies (CNES) has also opened its archive free for public with its Spot World Heritage (SWH) program. Users can now access Spot 1-5 images from 1986 to 2015 with the spatial resolution 2.5 to 20 m depending on the generation of the satellites [11].

Due to the advances in free data policy of space organizations, image archives going as far back as 1972 can be retrieved and processed. Developing countries can thus only focus to automatization, extract, interpret information, carry out retrospective analyses about LULC change and develop long-term plans for effective LA.

These advances in RS have made easier the performing of high-level tasks like classification to extract needed reliable information. There are available many strong concepts land cover features extraction and processing algorithms. Though classifiers may not use same data format, the outcomes precisions depend on the strength of the classifier. The classification results can be observed and correctly interpreted with help of user interference. For a better analysis, classification results can be compared one to another to point out the mis-correctly classified groups and the weakness of algorithms [12]. One should always use strong algorithms for specific projects where accuracy demands are high. This explains the usage of Machine Learning algorithms such as Random Forest (RF), Support Vector Machine (SVM) and the Neural Network (NN) learner, Deep Learning For Java (DL4J) in this study. These classifiers should boost considerably the classification accuracies and precisions with their sharpened analysis processes.

In RS and GIS environments, these methods can be useful for locations of restricted budgets. When it comes to obtaining a solid result in terms of accuracy and precision with low-cost materials. The challenge that arises lays in the user friendliness of different algorithms.

## 1.1 Problem Statement

LULC features classification-based analysis in RS, helps into giving sense to learning what has been taking place into a given area. Using robust clustering algorithms always boost the accuracy level of classification. How to select which algorithm is right for the data clustering purpose? Which one to apply and when to apply it? To obtain what result? These are generally questions that need a reply

before starting any classification project. A fit for purpose clustering method-based algorithm selection is always a good idea to be suggested. How can one use this method when the classification concerns LULC features or monitoring environment and urban changes? What about the concerned location for the classification? What if sporadic constructions are abundant and constitute the core of urbanization? Processing this data of Guinea, Forecariah region will be performed in this study, outcomes will be examined in the following chapters.

## 1.2 Research Objectives

Major objectives of this thesis are, first to use DL beside ML algorithms in RS LULC classification process. Analysis of the outcomes obtained from the applied classifications methods.

Reliability check of the outcomes.

To perform this study, some small tasks where to be handled.

- Sentinel data collection,
- Area of Interest extraction from the country data,
- Training and Test data extraction,
- Data manipulation,
- Data analysis and interpretation.

## 1.3 Thesis Outline

Chapter 1 presents the introduction, problem statement and thesis objectives. This part gives feedback on LULC classification. This can help public policy makers when it comes to looking for open data and tools for building their decision taking systems. Especially in LA, environment protecting and while conducting high engineering projects.

Chapter 2 talks about the different classification methods (unsupervised&supervised) applied in this work. It briefly describes the used classifiers that are K-means Clusters (K), Random Forest (RF),

Support Vector Machine (SVM) and Deep Learning For Java (DL4J). Mathematical formulas lying under all these methods are stated in.

Chapter 3 the application chapter, it introduce the data, study area, analyzes processes of data and results. It covers the discussion part as well. There advantages and disadvantages of all the applied methods are exposed to facilitate the comprehension. How LULC features monitoring can help in the build-up process of strong decision-making systems? How to keep under control all urban and environmental changes through LULC features classification?

Chapter 4 summarizes the mains conclusions and recommendations of the thesis.

# Chapter 2

## Land Use Land Cover Classification Methods

Mapping LULC has been gaining importance because the features on the earth's surface change within time. Temporal change at a location and its proximity gives clues about the conversion of the land use types. Monitoring the change asks for high accurate satellite data of the location at a given period. Processing and analyzing RS data will furnish imperative information for future critical decision-making processes such as efficient land administration, climate change issues. When it comes to locations where financial resources are limited, open access data sources can be used if the (free) processing methods are sufficient to yield strong results based on which future decisions will be adopted.

### 2.1 Land Use Land Cover Classification Systems

Since LULC classification is defined as a process of recognizing, clustering, categorizing and grouping all the ground features of a given location based on several criteria. In algorithms, are entered all the clustering criteria. Based on them, classification is carried out. In the computational processes, the changes between classifiers lie in both the data setting up processes and the entering criteria. Some classifiers use training and test set data (supervised classification) and some instead use iteration number (unsupervised classification). The classification algorithms, based on those prescriptions, calculate and process the grouping of LULC features. Different classification methods may be used accordingly for multiple reasons. Within those reasons, the study area, the aim of the study, etc... More of all, it's important to state that: the classification accuracy and precision are variables of the applied methods. Therefore, the methodology to be applied should be consequent. Thus, more users implication in terms of vigilance is requested. Some field inspections and site visits can be needed and better working plans drawn accordingly. We carried out two (2) classification

methods with four (4) different algorithms within this study. One (1) Unsupervised Classification Method : K-means Clustering Analysis (K), and three (3) Supervised Classification Methods : Random Forest (RF), Support Vector Machine (SVM) and Deep Learning For Java (DL4J).

### 2.1.1 Unsupervised Classification

It is a method that classifies LULC features based on the pixel brightness index. This method permits the user to proceed the features grouping according to their pixel value reflectance. It contains different classification algorithms that proceed the computation based on the user interference limitation styles. After delimitating the region to be classified, the iteration number is set, if no a priori class number is known, with help of the thumb rule, the class number can be determined and set. Then, the computer program proceeds the sampling process of features recognition. This computation is all based on mathematical principles that vary from algorithm to algorithm. Unlike supervised methods programs, these algorithms do not receive any signal of guidance from the user. Unless the class number is known a priori then, it can be defined in the program. Otherwise, the classifier groups LULC features after detecting their very nature while processing the dataset.

#### 2.1.1.1 K-Means Clusters Analysis Classification Method (K)

This unsupervised classification method is one of the most common use clustering algorithm. it serves as a foundation to many sophisticated clustering techniques[13]. It calculates initial class means evenly distributed in the data space then iteratively clusters the pixels into the nearest class using a minimum distance technique. Each iteration recalculates class means and reclassifies pixels with respect to the new means. All pixels are classified to the nearest class unless a standard deviation or distance threshold is specified, in which case some pixels may be unclassified if they do not meet the selected criteria [14]. This process continues until the number of pixels in each class changes by less than the selected pixel change threshold. Since K is a very sensitive algorithm to the number of clusters, large K cluster number is always suggested for improving the classification homogeneity factor. Therefore, a priori knowledge about the grouping is always necessary to setting the number of clusters. Or else, a thumb rule is suggested where number of cluster (k) refers to the square root of the half of the feature (n) present in the data  $\{k = \text{sqrt}(n/2)\}$ . In the same order, elbow method can be used to increase homogeneity and decrease heterogeneity until all data is well classified. As

it's clear, K means clustering algorithm does not give the labels of classes. It returns a group of classes having features of similar patterns. Task to the user to label the groups returned by the analysis result [13].

## 2.1.2 Supervised Classification

This method is processed with a strong user implication theory. The idea behind it, is based on the user to detect and defines samples in the data according to his knowledge of the study area. Every feature sample defined in the Area Of Interest (AOI), constitute one feature class group. By this process, the algorithm is then furnished in terms of training and tests datasets. These training and test samples constitute the two (2) parts of the defined AOI. The user knowledge is utilized to set the boundaries between features which corresponds to assigning and labeling pixel brightness values to features. Most importantly, the user defines the number of classes to be present in the study based on the ground reality [15].

### 2.2.2.1 Random Forest Classification Method

RF classifier consists of combining many individual decision trees that operates together. Literally, it is an ensemble learning method that constructs a multitude of decision trees as a training set. Based on the wisdom of crowds, the better outcomes are obtained when the operating groups of trainings sets (decision trees) are uncorrelated. Because trees protect each other from their individual errors. It can therefore be defined as a classification algorithm consisting of many decision trees. It proceeds by bagging and feature randomness while building every singular tree. Then, it creates an uncorrelated forest of trees whose prediction by committee is more accurate than the prediction of individual tree. RF is performed to solve many different problems beside classification, such as regression tasks [16].

It performs by leveraging the power of multiple DTs. Based on that process, the prediction of the decision is obtained by taking the average mean of the output from various trees. RF algorithm is a sum of the output of several randomly created DTs. From this method, a very precise and accurate decision is accessed as final output. Beside that, increasing the number of trees, implies the increase of accuracy of the outcome. RF, boosts the precision of DT by reducing overfitting of datasets. Therefore, a better understanding of DT becomes a must to explain the process that takes place in



the RF classification. They (DT) are building blocks for RF algorithms for which, they create a structure like a tree. DT, has 3 components such as:

- \* Root Nodes = {Main Attribute branch}
- \* Decision Nodes = {First Attributes branch}
- \* Leaf Nodes = {Second to last Attributes branch under Decision Node}

The training data is divided into several attribute branches. The attribute branches too explode into multiple branches. This sequence continues taking place until a leaf node is obtained. The leaf node cannot be segregated anymore into new branch. All these processes take place under the root node to which they are related [17]. It's shown in Figure 2.1.

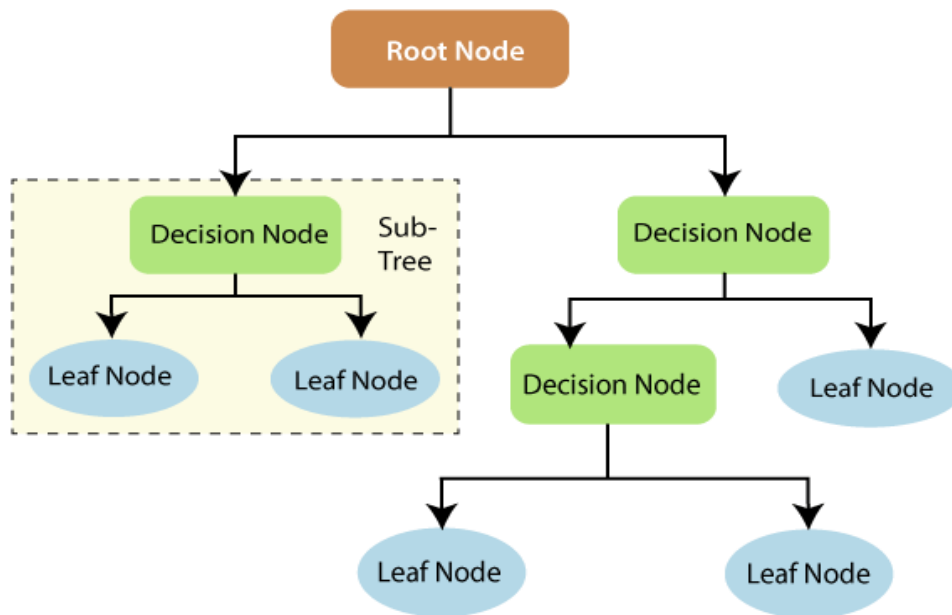


Figure 2.1 Decision Tree (image gathered from Tutorial examples)

### Implementation Formula in Scikit-learn (formula obtained from DT tutorial homepage:

<https://towardsdatascience.com/the-mathematics-of-decision-trees-random-forest-and-feature-importance-in-scikit-learn-and-spark-f2861df67e3>)

\*For each decision tree, Scikit-learn calculates a nodes importance using Gini Importance, assuming only two child nodes (binary tree):

$$NI_j = W_j C_j - ((W_{left(j)} * C_{left(j)}) - (W_{right(j)} * C_{right(j)}))$$

- $NI_{sub(j)}$  = the importance of node j
- $W_{sub(j)}$  = weighted number of samples reaching node j
- $C_{sub(j)}$  = the impurity value of node j
- $left(j)$  = child node from left split on node j
- $right(j)$  = child node from right split on node j

\*The importance for each feature on a decision tree is then calculated as:

$$fi_i = \frac{\sum_{j: \text{no de } j \text{ split on feature } i} ni_j}{\sum_{k \in \text{all nodes}} nik}$$

- $fi_{sub(i)}$  = the importance of feature i
- $ni_{sub(j)}$  = the importance of node j

\*It is then normalized to a value between 0 and 1 by dividing by the sum of all feature importance values:

$$normfi_i = \frac{fi_i}{\sum_{j \in \text{all features}} fi_j}$$

\*The final feature importance, at the Random Forest level, is its average over all the trees. The sum of the feature's importance value on each trees is calculated and divided by the total number of trees:

$$RFfi_i = \frac{\sum_{j \in \text{all trees}} normfi_{ij}}{T}$$

- $RFfi_{sub(i)}$  = the importance of feature i calculated from all trees in the Random Forest model
- $normfi_{sub(ij)}$  = the normalized feature importance for i in tree j
- $T$  = total number of trees

### 2.1.2.2 Support Vector Machine Classification Method

SVM is a ML algorithm in which both classification and regression challenges can be solved. By plotting every single data item in an n-dimensional space (n is data features number), the value of each feature is considerate as the value of a particular coordinate. It segregates the data classes by using hyperplane (line). This works best on complex datasets due to its strength and power in building ML models. The SVM classification consists of finding the hyperplane location for separating

different data types. It has a tool that ignores outliers and finds hyperplane with the maximum margin (see Figure 2.2).

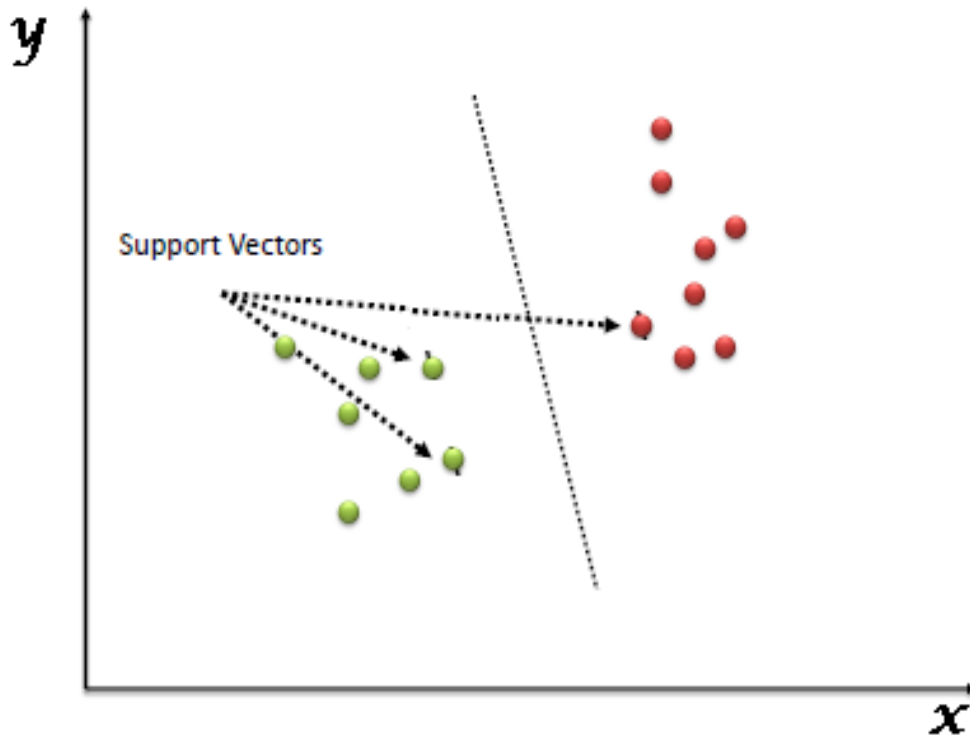


Figure 2.2 Support Vector Machine (image gathered from Tutorial examples)

Therefore, it is a robust classification algorithm to outliers. For more complex cases, it uses kernel trick that takes low dimensional input space and converts it into a higher dimensional space. In other words, it transforms not separable problems into separable ones. This process is mostly used in nonlinear classification projects [18].

**Radial basis function kernel (RBF)/ Gaussian Kernel** (formula obtained from SVM tutorial homepage: <https://medium.com/@zxr.nju/what-is-the-kernel-trick-why-is-it-important-98a98db0961d> )

\*Gaussian RBF(Radial Basis Function) is another popular Kernel method used in SVM models. RBF kernel is a function whose value depends on the distance from the origin or from some point. Gaussian Kernel is of the following format;

$$K(X1, X2) = \text{exponent}(-y||X1 - X2||^2)$$

$$\text{Euclidean}(X1, X2) = ||X1 - X2||$$

Using the distance in the original space we calculate the dot product (similarity) of X1 & X2.

*Note: similarity is the angular distance between two points.*

### 2.1.2.3 Deep Learning Classification Method

DL, known as an AI function, with its NN process, imitates the human brain when it comes to analyzing data. It is very broad and complex to navigate. For that reason, one needs a global and non-ambiguous understanding when it comes to ruining a project under its algorithms. DL gives the possibility to its users to apply both supervised and unsupervised methods based on the needs

and the problems. Its particularity is the usage of intuitive concepts in solution searching beside all the mathematical scripts. It combines various systems via NN. Mainly , it is applied in the object detection, speech recognition, language translation and for decision making processes in various

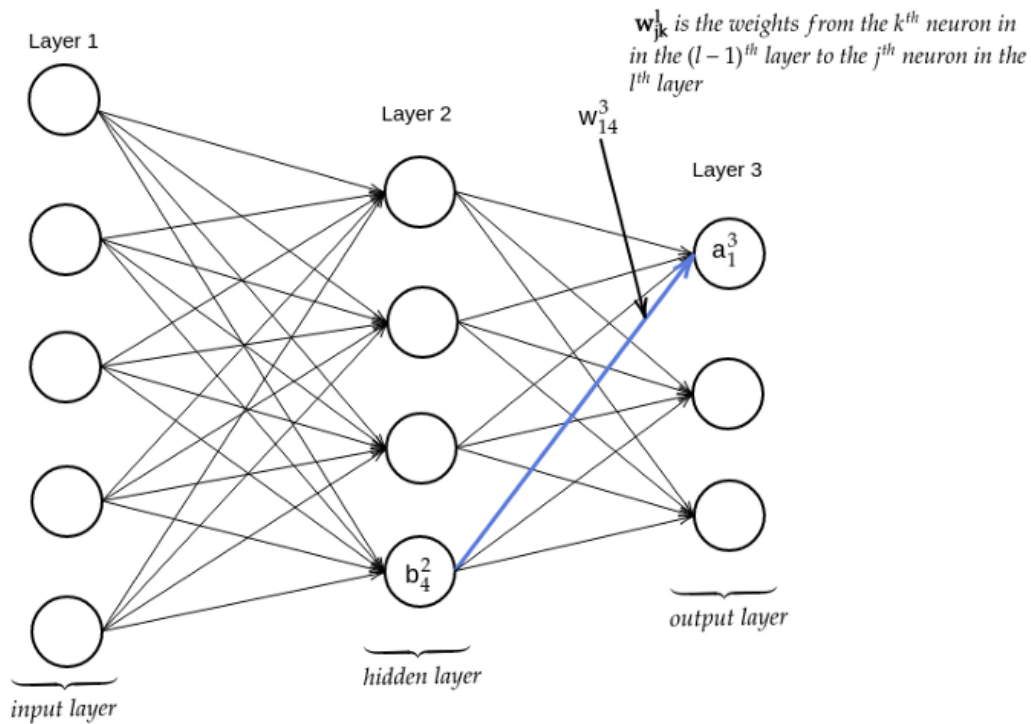


Figure 2.3 Deep Learning Neural Network Processing Step (image gathered from Tutorial examples)

application domains. It utilizes frameworks which allow one to develop tools able to offer better attractions while simplifying hard programming issues. Within those frameworks, there's Tensorflow, Caffe2, DL4J, and so forth [19].

# Chapter 3

## Application of Land Use Land Cover Classification in Forecariah

### 3.1 Study Area and Data: Forecariah, Guinea

#### 3.1.1 Study Area: Forecariah, Guinea

For this study, the concerned area was selected in Guinea (West African Country) with Area of 245,857 sqkm. Guinea lies between 7.19° and 12.7° N latitudes and between 7.6° and 15.1° W longitudes (Figure 1.1). It includes the Los Islands, which are a group of island located in the west of the capital city Conakry. In Guinea, the seacoast is marshy and is about 274 km long. The interior rises to hilly and plateau regions. Mount Nimba (1,752 m), the highest point, is located near the borders with Ivory Coast and Liberia [20]



Figure 3.1: Map of Guinea ([countries.mapsofworld.com](http://countries.mapsofworld.com))

Forecariah is one of the prefectures in Guinea, with an area of 4384 sqkm, whose capital city is Forecariah-Centre (see Figure 1.2). The prefecture has a population of over 240,000 residents. The region was selected as a case study because it lies close to Conakry, the capital city of Guinea. Conakry has more than one fifth (1/5) of the country's population. Recognized as the country's trade and political center, it attracts people from other cities. Previous urbanization studies indicated that Conakry's rapid urban growth can be explained by its mass migration attraction from the rural areas to the urban areas for socioeconomic and physical factors. The urban population was estimated at nearly 3.7 million, which is 34.7% of the country population [21,22].

Forecariah suffers the consequences of Conakry's rapid expansion which characterizes the systematic conversion of rural farming lands into urban areas. In the other hand, the mine exploitation activities that have been started in the region accelerates the environmental change (pollution) within Forecariah. The area has a tropical climate with only two seasons: the sunny season and the raining

season with 3784 mm/year of precipitation [23]. In 2018, Guinea had an estimated population of 13.05 millions, which ranks 74th in the world. The population grew from 8.7 million in 2000 to its current state, which corresponds to 3.3% annual increase. Although the country has rich mineral reserves of bauxite, iron, gold, diamond, its mild climate enables the production of banana, cacao, cafe, pineapple, mango, orange, cotton, papaya, cashew, rice, potatoes, etc... In Forecariah, farming and fishing are most practiced activities. Due to start-up of mining activities and the rapid expansion of Conakry to its surrounding cities, sporadic land occupation and management have been observed. As a direct consequence, living conditions are impacted. Carrying out an exhaustive diagnostic of that fact, should be a must for public decisions and policy makers. For us, performing such study consists of giving a trial to identification and solution suggestions to that reality with help of open access data and tools. RS and GIS, with help of new technology approach-based data and tools availability, can be used for such analysis.

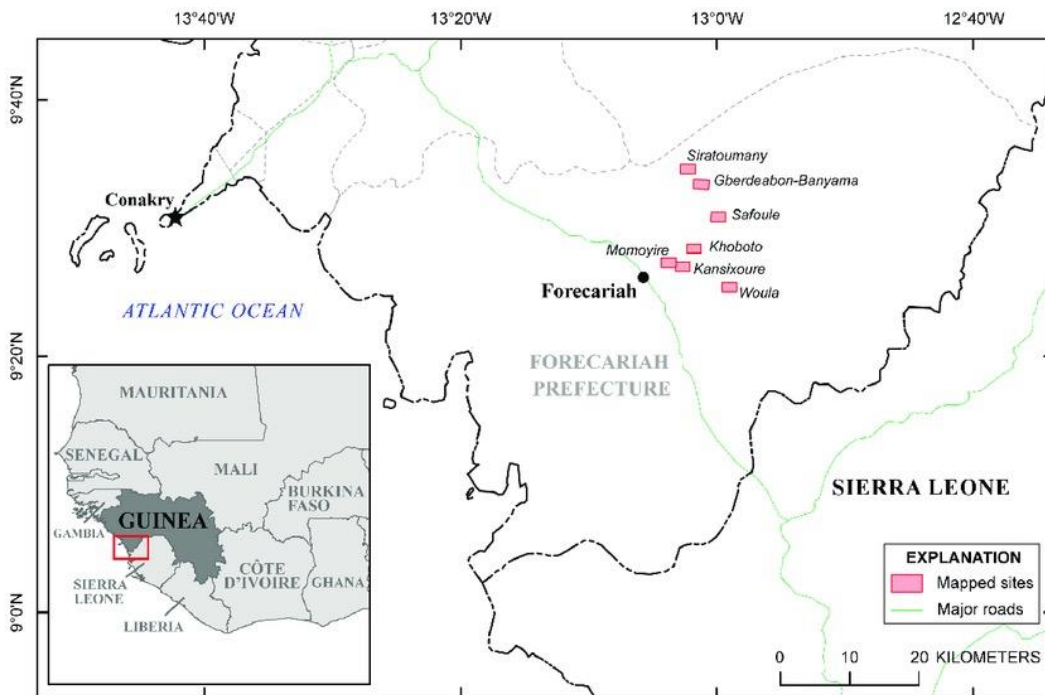


Figure 3.2: Map of Forecariah (countries.mapsofworld.com)



### 3.1.2 Data

Satellites, with their constellation systems, help into providing precise data for multiples RS projects analyze purposes. Within those satellite data sources, the US Landsat system has readily been available since 70s and has been providing users with 30 m resolution. A newer system initiated by the Copernicus Program of the EU, Sentinel 2, delivers images with 10 m resolution in visible and near-infrared (NIR) region since 2015. The users can download images of 5-day-temporal resolution acquired by the twin satellites (Sentinel 2A/2B). This high temporal and spatial resolution can be used for monitoring developing areas. Such a developing region, Forecariah, has been selected for LULC mapping. Copernicus Open Access Hub (<https://scihub.copernicus.eu/>), Sentinel-2A image was downloaded and analyzed within this study. Acquisition date of the atmospherically, geometrically and radiometrically corrected level 2A image is February 19, 2019. Sentinel-2A image provides 10m spatial resolution in four (4) of its 13 bands. We extracted these four bands including visible and NIR, b2, b3, b4, b8. (see figure 2.1).

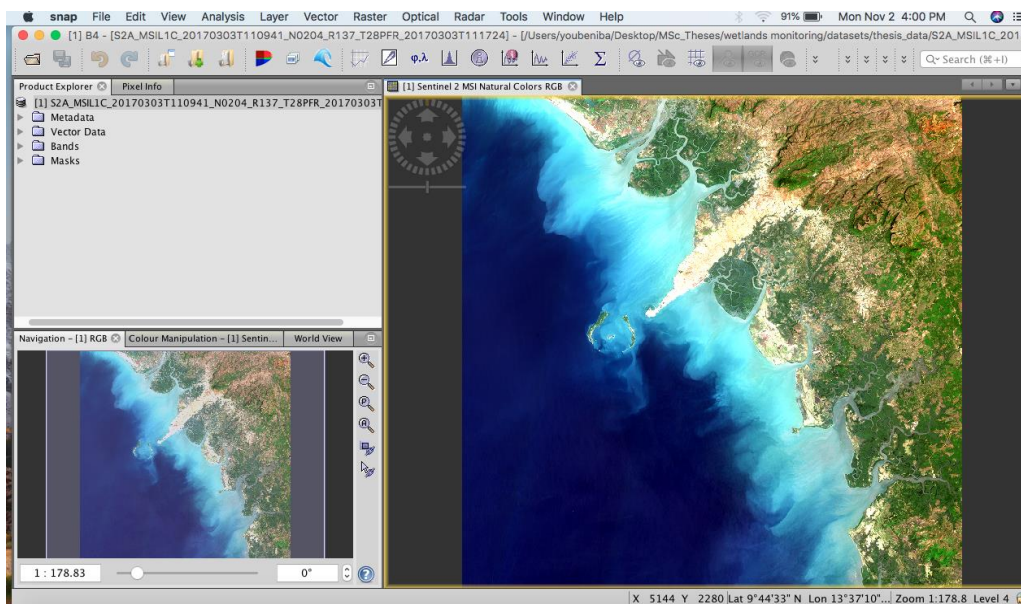


Figure 3.3: Sentinel 2A image from February 19, 2019 displayed in SNAP

From the 10m resolution image data, we spatially created subset of the image under the Sentinel Application Platform (SNAP 6.0). In order to have under control to focusing for the study purpose, only our study area. As it can be seen in Figure 3.4. and the image data 3.5.

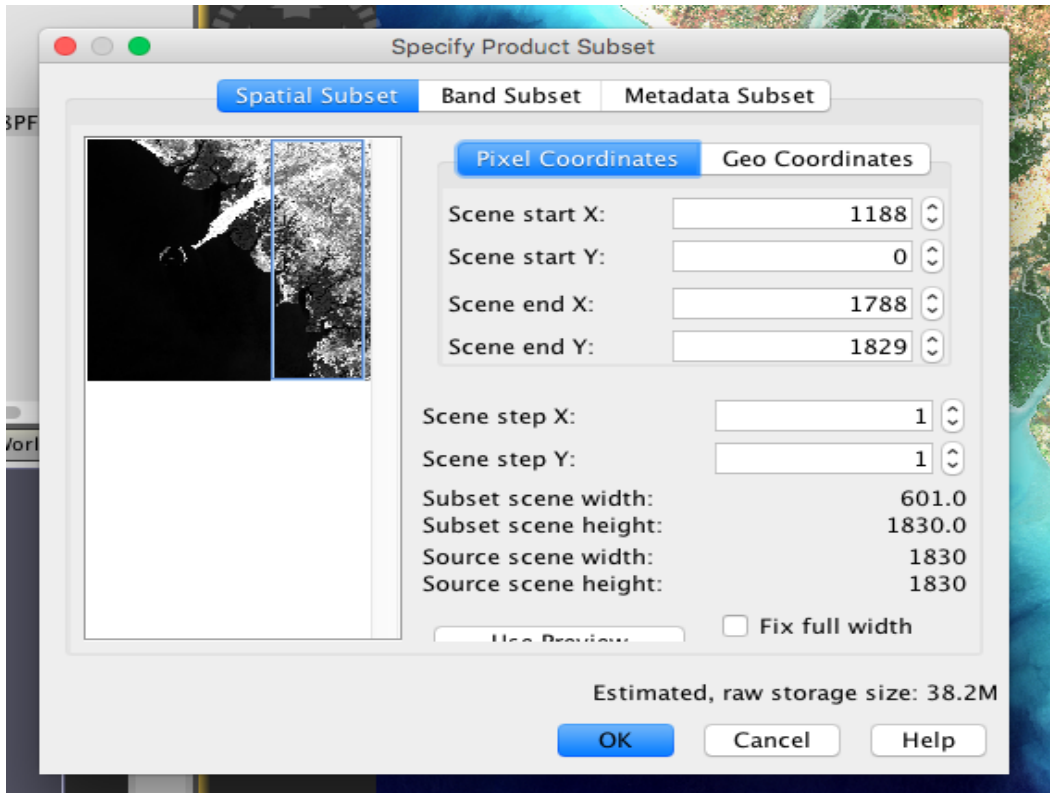


Figure 3.4: Subsetting process of our raster data under SNAP environment



Figure 3.5: Subsetted raster data of Forecariah extracted from the atomic data of Sentinel-2A

This pre-processing part was performed under Sentinel Application Platform (SNAP 6.0). Since the acquired data for this study, was an atmospherically, geometrically and radiometrically corrected, Sentinel level 2A, ortho-rectified data. No calibrations were needed to be applied to the data. See Figures 3.6 and 3.7.

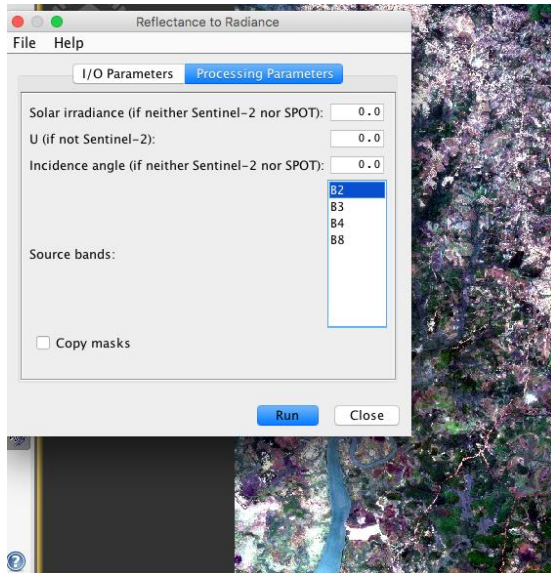


Figure 3.6: No Radiometric Correction was needed

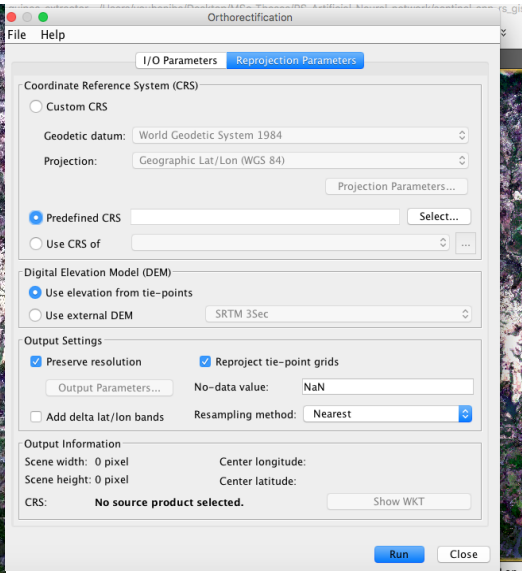


Figure 3.7: No Geometric Correction was needed

Train and test polygons were collected in the image by using Google Earth (<https://googlemaps.com>), which hosts very high-resolution images of earth surface features. Land cover class names were adopted from the World First Global Land Cover Datasets (WFGLCD) at 30m Resolution. This open access high resolution map of the earth features was donated to the United Nations (UN) by China in 2014 (<http://www.globallandcover.com>) [24]. This classification of Forecariah, contains classes such as: Water, Wetland, Agricultural Area, Vegetation Area, Bare-land and Built-up (see Figure 3.8).

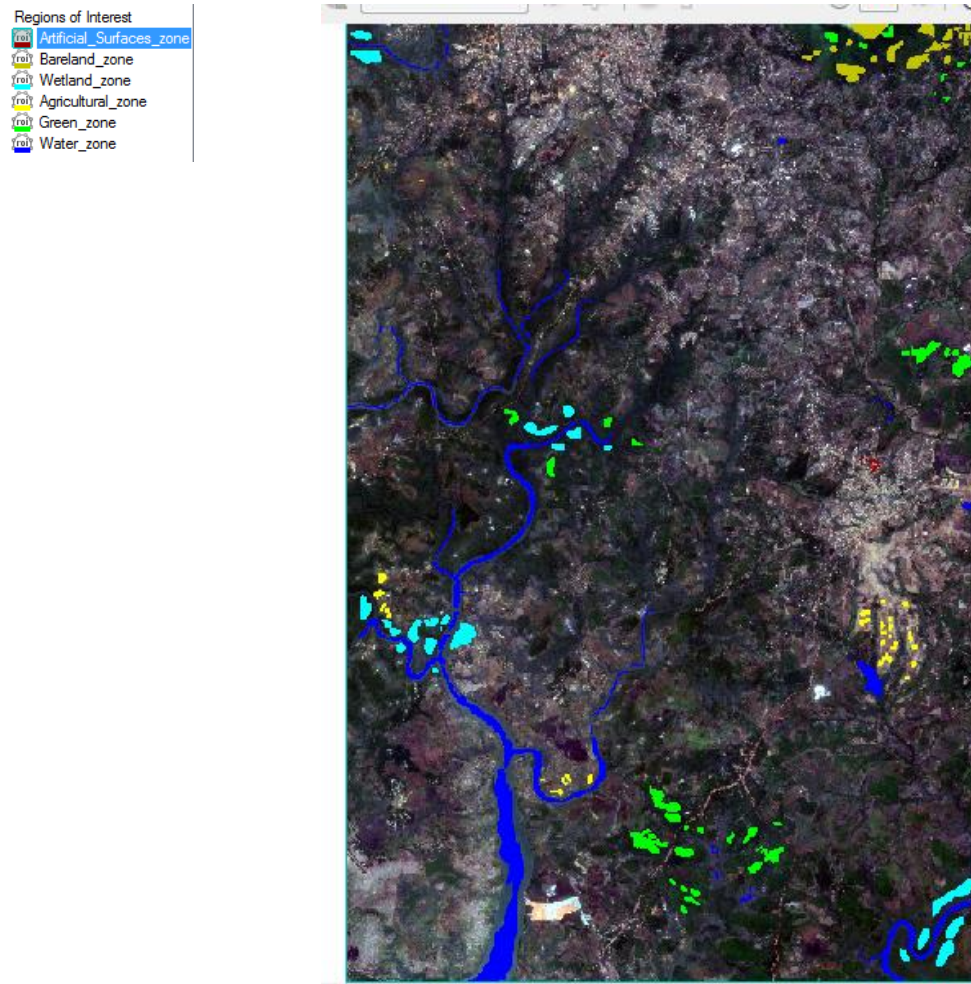


Figure 3.8: Region of interest

Our analysis consisted of applying both the unsupervised and supervised classification methods. We carried out four (4) classification algorithms within this study: one (1) Unsupervised Classification Method, K-means Clustering Analysis (K), and three (3) Supervised Classification Methods, Random Forest (RF), Support Vector Machine (SVM) and Deep Learning For Java (DL4J).

### 3.1.2.1 K-Means Clusters Analysis Classification Method (K)

In the processing steps, since from the World First Global Land Cover Datasets, there are 6 classes available in the Forecariah region, we fed the number of clusters to 6 and get the iteration number set to 30. Then, based on the pixel brightness index, the program itself computed the result and the

accuracy of it. The overall accuracy of 78.13% was obtained as well as the classification image outcome. The classification map and table are shown in Figure 3.12 and table 3.1.

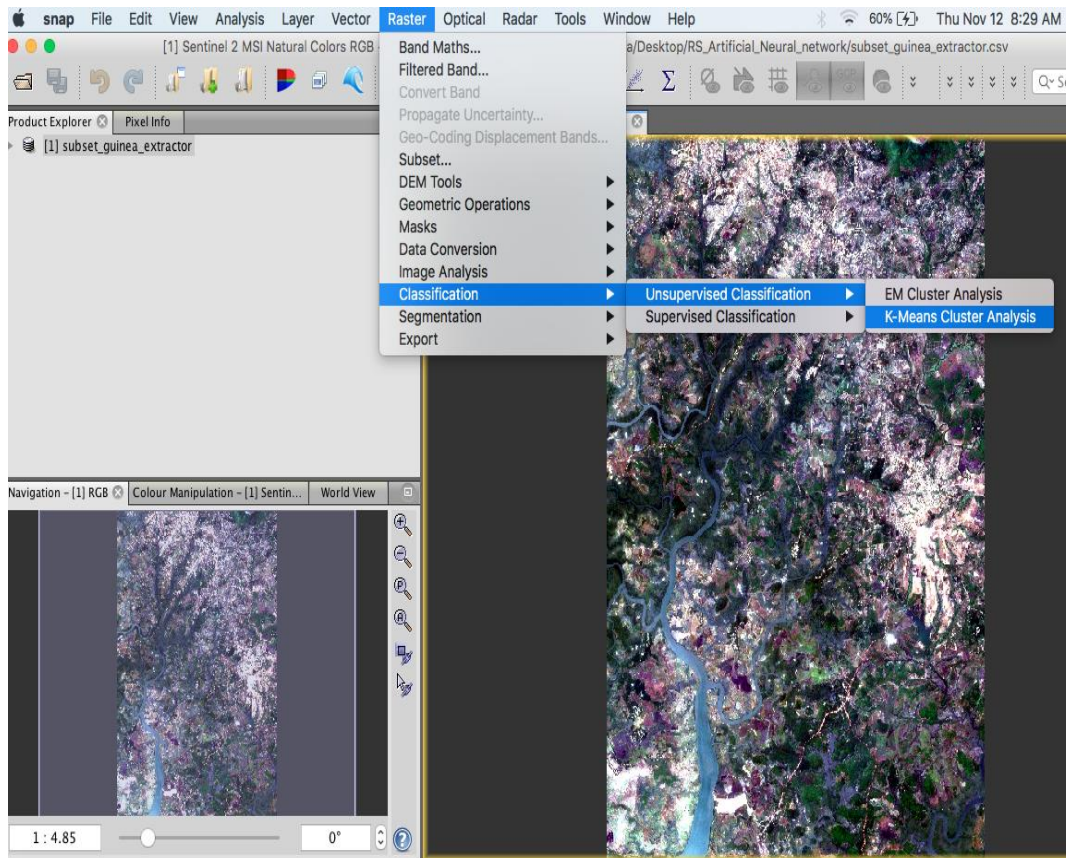


Figure 3.9: K-Means processing selection

To be sure that, there are 6 classes available in the Forecariah region, we proceed the processing of the dataset by using the default (maximum class number) cluster value which is 14 in the K case from ESNAP. As a result, we observed that the data was clustered into 6 groups. Therefore, the number of classes attributed to our study area by the World First Global Land Cover Datasets was set to be conciliated see the Figures 3.10 & 3.11.

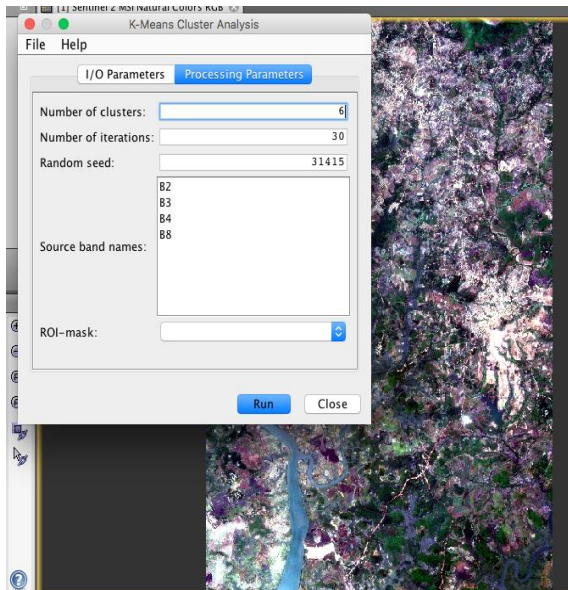


Figure 3.10: Number of clusters and iteration setting

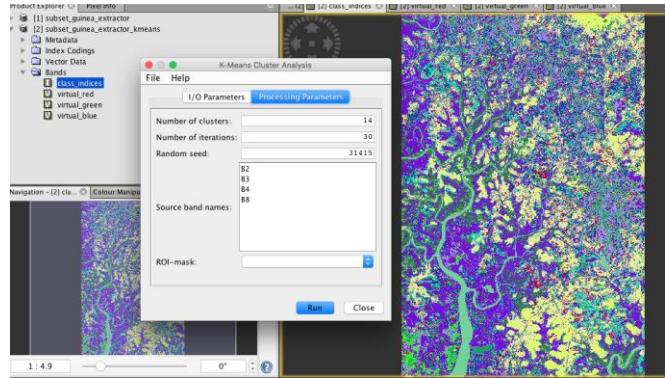


Figure 3.11: Using default number of clusters and iterations

Table 3.1: K-Mean Cluster Analysis Results

	Measures	Estimate [%]
	Overall accuracy	78.13
	Kappa accuracy	73.12
	User's accuracy [%]	Producer's accuracy [%]
Map class	Estimate	Estimate
1 water	88.3	88.3
2 wetland	84.2	82.9
3 agriculture	89.2	89.0
4 vegetation	87.7	85.1
5 bareland	64.0	74.3
6 built-up	72.8	80.3



Figure 3.12: K-Means classification map



### 3.1.2.2 Random Forest Classification Method

In this specific case, we classified our test site by using RF based on DT method. RF generated many DTs by using randomly selected training and testing vectors so that it produces the best vote for the decision. EnMAP-Box (EnMAP-Box 3.7) was used.

Environmental Mapping and Analysis Program (EnMAP) together with its toolbox and application programming interface (EnMAP-Box) is a pure German developed invention technology. EnMAP is a hyper-spectral satellite mission. It aims to monitor and characterize the Earth's surface features and environment on a global scale. Its toolbox, EnMAP-Box is mainly a graphical user interface developed for visualizing raster and vector data types. It is a python plug-in for Quantum Geographic Information Systems (QGIS). Designed to process and visualize hyper spectral RS data. With an ability to compute and model the earth ecosystem dynamic processes, EnMAP extracts and gather various and valuable information about the earth's surface features [25].

Since EnMAP-Box integrates ML algorithms into the data classification and regression analysis in the QGIS environment, we used RF library available under scikit-learn which is a ML platform working under Python environment (<https://scikit-learn.org>).

Within this process, all the training and test raster dataset previously delimited under SNAP (AOI), were converted as shape file under QGIS so that it can be used in the classification workflow under Scikit-learn RF processing (see Figure 3.13).

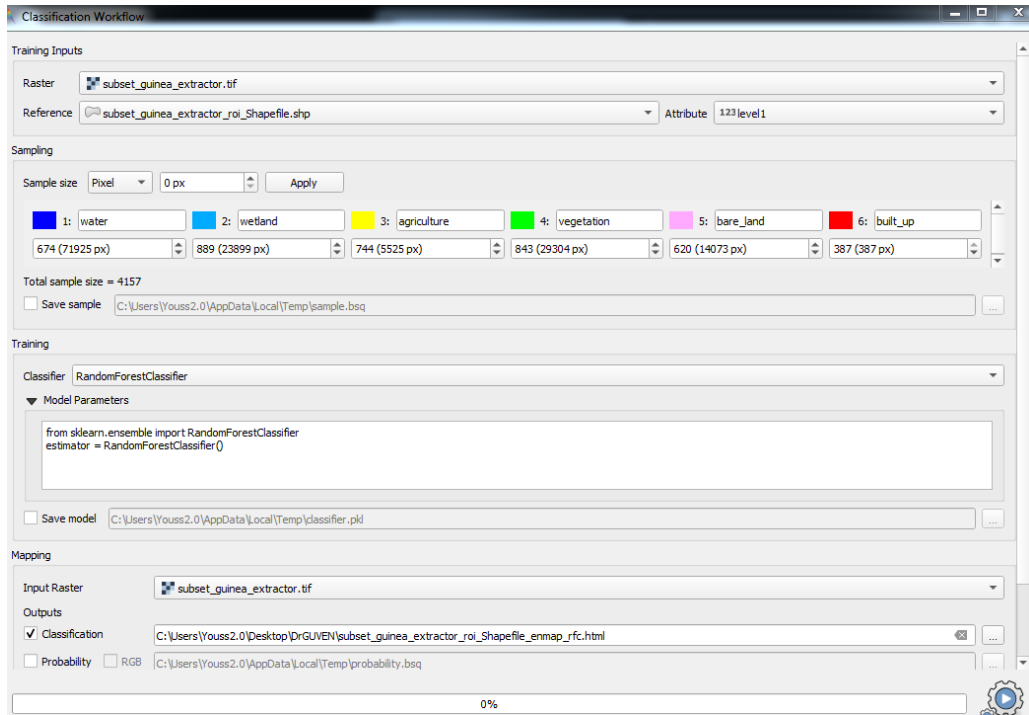


Figure 3.13: RF classification method settings

We fed the number of decision trees as 100 and the program determined the number of randomly selected features as the square root of the number of features. The accuracy of each classification was assessed by using 10-fold cross validation. The 10-fold cross validation divided the ground truth data into ten (10) equal size subgroups and each 10% subset is selected iteratively as validation data set. Ten (10) classifications were then carried out with the remaining %90 of the entire ground truth data. We finally computed thus accuracies ten (10) times. The accuracies were obtained and their overall was 83.35%. The classification image outcome was also obtained. Table 3.2 presents the overall and class-wise accuracies, the classification result is given in Figure 3.14.

Table 3.2: RF cross-validation results

ACCURACIES						
Measures	Estimate [%]		95% confidence interval [%]			
Overall accuracy	83.3		82.0		84.5	
Kappa accuracy	79.8		78.4		81.2	
Map class	User's accuracy [%]			Producer's accuracy [%]		
	Estimate	95% interval		Estimate	95% interval	
1 water	99.0	98.6	99.3	98.5	97.8	99.3
2 wetland	80.7	79.5	81.9	80.8	78.5	83.0
3 agriculture	83.1	82.1	84.2	86.8	84.6	89.1
4 vegetation	81.4	80.3	82.6	82.8	80.5	85.1
5 bareland	74.7	73.3	76.0	74.3	68.2	74.4
6 built-up	80.27	78.99	81.55	76.74	73.12	80.36

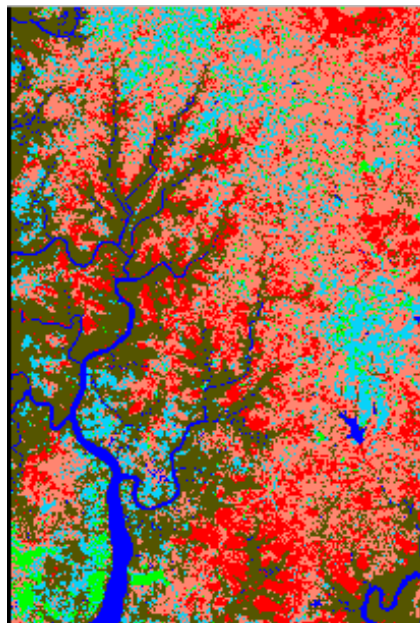


Figure 3.14: RF classification map

### 3.1.2.3 Support Vector Machine Classification Method

In this supervised classification method, we applied the non-linear algorithm which finds the best hyperplane maximizing the margin between the classes. SVM classifier under EnMAP-Box was utilized. We used SVM library available under scikit-learn. Within this process, the already converted training and test raster dataset into shape files was processed. As kernel, we used the Gaussian radial basis function whose parameters were adopted as default values broadcasted in the classification workflow of the program.

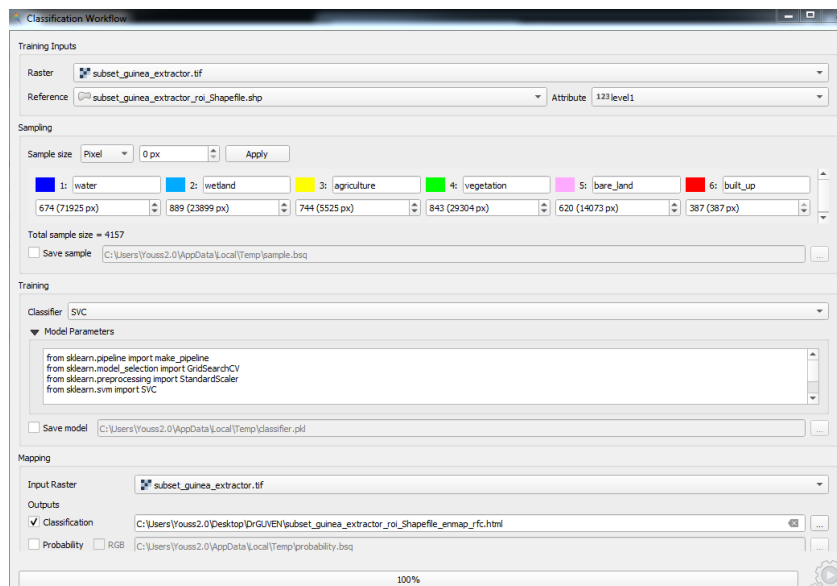


Figure 3.15 SVM classification method settings

We assessed the accuracy of each classification by using again 10-fold cross validation, where the ground truth data were divided into ten (10) equal size subgroups and each 10% subset is selected iteratively as validation data set. Ten (10) classifications were then carried out with the remaining %90 of the entire ground truth data. We finally computed thus accuracies ten (10) times. The accuracies were obtained and their overall was then 86.22%. See the cross-validation outcome and the classification image (see Table 3.3 and Figure 3.16).

Table 3.3: SVM cross-validation results

Measures	Estimate [%]	95% confidence interval [%]	
Overall accuracy	86.2	85.0	87.2
Kappa accuracy	83.2	82.0	84.5

Map class	User's accuracy [%]			Producer's accuracy [%]		
	Estimate	95% interval		Estimate	95% interval	
1 water	99.4	99.1	99.7	99.3	98.7	99.8
2 wetland	79.8	78.8	80.9	86.3	84.2	88.4
3 agriculture	85.0	84.1	85.9	91.4	89.3	99.5
4 vegetation	86.5	85.6	87.4	89.8	87.7	91.9
5 bareland	82.5	81.1	84.0	67.1	64.0	70.1
6 built-up	85.7	84.5	87.1	76.2	72.8	79.6

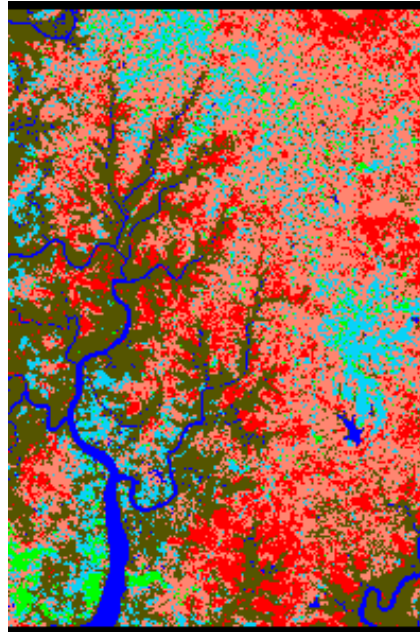


Figure 3.16 SVM classification map

### 3.1.2.4 Deep Learning For Java Classification Method

DL has a strong functions that can learn without any human supervision. Even from unstructured, unlabeled, and scratched data, it can learn by simple drawings. It uses methods based on a NN classifiers. The underlying problem is formulated as a network architecture whose output layer defines a loss function needed for learning. The output units can be formulated as probabilistic predictions, and if these predictions are parameterized, one can simply define the loss to be the negative log-likelihood under the model [26].

We carried out this DL classification by using Eclipse Deep Learning For Java (DL4J) algorithm under Waikato Environment for Knowledge Analysis (Weka) (see Figure 3.17).

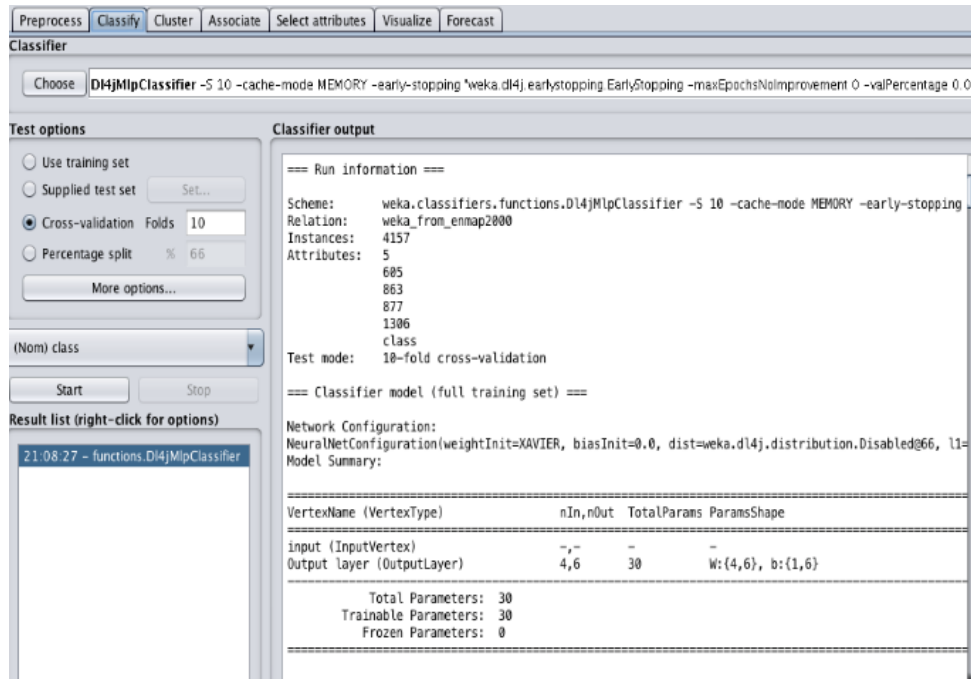


Figure 3.17 Weka DL environment

While EnMAP-Box accepts training and test pixels as shape file, Weka doesn't. It accepts data only if text format. It recognizes inputs with Attribute Relational File Format extension (.arff). Weka is a free software developed by the Waikato University of New-Zealand. In it, there are algorithms and tools for both data analysis and predictive modeling visualization. With an easy-to-use graphical interface, accessible to all, its inner functions manipulation are friendly usable. Weka supports many standards data mining tasks, especially data features selection, pre-processing, regression, classification, clustering, and visualization [27].

Since Weka does not recognize the image file format, we converted our image data to text (csv) file format under SNAP. As a result, we obtained more than six (6) millions rows. Based on a small amount of randomly extracted rows from our data, used as samples, we created a model. Within the parameters we set the number of epochs to 10 and the batch size to 100. This model was run for the LULC classification prediction of our dataset. After loading the model from the supply test set in the

pop up menu, we reevaluated the newly loaded dataset to be predicted. Note that in the dataset to be analyzed, the column to be evaluated, classified and predicted should contain only question marks (?). This process is less time consuming and accurate (model based analysis).

As we processed in the other supervised classification methods, the accuracy of the DL4J classification was assessed by using 10-fold cross validation. Because in 10-fold cross validation, the ground truth data are divided into ten (10) equal size subgroups and each 10% subset is selected iteratively as validation dataset. Ten (10) classifications were then carried out with the remaining %90 of the entire ground truth data. The accuracies are finally computed thus ten (10) times. The accuracies and their overall was obtained as 91.32%. The classification outcome and the image were presented (see Table 3.4 and Figure 3.18).

Table 3.4: DL4J cross-validation results

	Measures	Estimate [%]
	Overall accuracy	91.3
	Kappa accuracy	89.4
	User's accuracy [%]	Producer's accuracy [%]
Map class	Estimate	Estimate
1 water	94.5	97.5
2 wetland	88.3	96.9
3 agriculture	90.1	94.1
4 vegetation	97.9	96.1
5 bareland	88.0	73.4
6 built-up	87.4	80.9



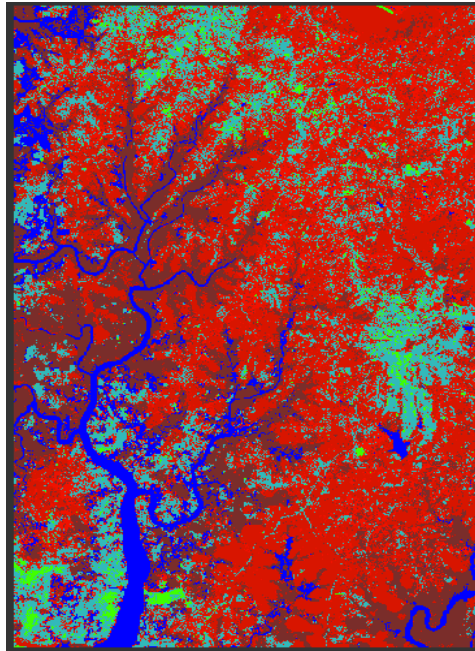


Figure 3.18: DL4J classification map

## 3.1 Pros and Cons

### 3.1.1 Results

We assessed the accuracy of each classification by using 10-fold cross validation. Which divides the ground truth data into ten (10) equal size subgroups and each 10% subset is selected iteratively as validation data set. All the supervised classification methods were processed by applying this validation method. Except the unsupervised classification case, where the class number check was carried out. It was found corresponding to the class number (6) given by the World First Global Land Cover Datasets to our study area. With respect to that class number of 6 features, we computed thus accuracies ten (10) times for supervised methods and the iteration number was set to 30 for the unsupervised method. The overall accuracies were obtained as 78.13%, 83.34%, 86.22% and 91.32% for K, RF, SVM and DL4J, respectively (Figs 3.12, 3.14, 3.16, 3.18). SVM

gave slightly better results than the RF and much better compared to K. RF and K took less computational demand in terms of time than the other two (2) algorithms. Where DL4J computational time was far away the highest. On a Pentium 5 machine with 16GB RAM capacity, K performing outcome was obtained in 5min while RF performed the computation in half a day (720min). Computation times for SVM and DL4J were both approximately one full day (1440min) and three full days (4320min), respectively. See classification results in table 3.5.

Table 3.5 classification results

ALGORITHM		ACCURACIES
DL4J	OVERALL ACC	91.32%
	KAPPA ACC	89.44%
SVM	OVERALL ACC	86.22%
	KAPPA ACC	83.22%
RF	OVERALL ACC	83.35%
	KAPPA ACC	79.78%
K	OVERALL ACC	78.13%
	KAPPA ACC	73.12%

### 3.1.2 Discussions

The best accurately classified class was observed as water class, whose brightness values differ from other classes. We obtained the bare-land class as the least accurate class in RF and DL4J but not in SVM and in K. Bare-land and wetland are two (2) related feature classes in the reality on the earth surface. For that reason, they can get to mix up slightly one into each other. The location of the pixels belonging to these classes lie within their transition zone because.

When we take a look to the spatial classifications outputs of all the four (4) classifiers, there's a slight difference between features occupied spaces in terms percentage. The variation rate lies between 0 to 3 (0%-3%) for most of the feature classes. Only between the SVM and K methods, we noticed a huge variation between vegetation and bare-land feature classes. In the SVM method, vegetation area covers 29.4% where in the K method, it corresponds to 20.8%. In the other hand, in SVM method, bare-land corresponds to 19.8% while in the K method, it corresponds to 26.7%. This shows a difference of 8.6% within the 2 vegetation classes and 6.9% between the 2 bare-land classes. This is explained by the classification of newly planted zone in

vegetation area by vegetation in the SVM method while in the K method, all those areas are depicted as bare-land (not vegetation) for having same characteristics as bare-land. See table 3.6.

Table 3.6 CLASS WISE ACCURACIES

Map class	User's accuracy [%]				Producer's accuracy [%]				Spatial Surface [%]				Area [sqkm]			
	DL4 J	SV M	DT	K	DL4 J	SV M	DT	K	DL 4J	SV M	DT	K	DL4 J	SV M	DT	K
1 water	94.5	99.4	99.0	88.3	97.5	99.3	98.5	88.3	16.1	14.9	15.7	16.5	94.2	87.2	91.8	96.5
2 wetland	88.3	79.8	80.7	84.2	96.9	86.3	80.8	82.9	3.7	4.3	5.0	2.9	21.6	25.2	29.3	16.9
3 agriculture	90.1	85.0	83.1	89.2	94.1	91.4	86.8	89.0	11.3	12.1	12.0	14.2	66.1	70.8	70.2	83.1
4 vegetation	97.9	86.5	81.4	87.7	96.1	89.8	82.8	85.1	27.5	29.4	28.1	20.8	160.9	172.0	164.4	121.7
5 bareland	88.0	82.5	74.7	64.0	73.4	67.1	74.3	74.3	21.0	19.8	20.2	26.7	122.9	115.8	118.2	156.2
6 built-up	87.4	85.7	80.2	72.8	80.9	76.2	73.1	80.3	20.3	19.5	19.0	18.9	118.7	114.1	111.0	110.6

DL4J came up with the best result among the studied classifiers. The disadvantage of the method is, its high computational cost leading to longer processing time. For the ones, interested into good accuracy only, by using open source data and tools, this method can be advised. Despite its long computational time. Or else, by providing some other costly software available on the market, one can solve these problems with instead of expensive time, they will face expensive software price to afford. For that reason, we are planning to keep on testing some other DL algorithms on free platforms. Orfeo Toolbox which uses Tensor-flow for example can be used for other studies again in developing countries where financial resources are limited. In the future, for LULC, change detection analysis and environment monitoring issues, etc, one can use those freely available DL algorithms for the processing. Especially for the zones similar to Forecariah.

Where sporadic urbanizations are abundant and budget is limited. Because of the strength of the DL classifiers and the accuracies they provide to users. This process should continue until both good outcome results are obtained within a reasonable computational time. Public policy makers should always monitor the study area by using open source data and tools. This process may continue until a general change analysis map is created for the region. This map will then be kept updated at certain intervals to control all the illegal activities occurring in the jurisdiction. This will be a data to be utilized by policy makers to keep tracking all LULC issues. These maps can be used within a decision-based system which then will improve the LA, climate change monitoring capability, marine control and so forth by the authorities at low cost.

# Chapter 4

## Conclusion and Recommendations

Developing a strong system free of charge for developing countries, ready to use in the decision making process is a must. Especially, when it comes to monitoring LULC features, planning better urban LA and rural LC, controlling climate change and tracking LULC changes. To build such a system, one needs to carry out a simulation based on the usage of multiple classifiers. The method to be selected should be based on both the cost in terms of time and expense. In this study, we used open source data and tools for performing analysis and we obtained satisfying results based on the accuracies. Both supervised and unsupervised methods were used. Since it is a trial study, for building decision support system, both the supervised and unsupervised classifiers were used. RF, SVM and DL4J as supervised method were applied and K, as unsupervised method was applied. To finish, this study can be used as a basement for helping both public policy makers and other private engineering project conductors when it comes to classifying LULC features in developing countries.

### 4.1 Conclusion

Satellite Sensors with high time resolution can provide valuable data and tools to monitor the changes occurring on land due to anthropogenic pressure. Sentinel 2 MSI is one such system delivering high resolution images with 10 m ground sampling distance and 5-day revisit frequency. Located in a close proximity to Conakry, where more than 1/5 of the country's population live. Sentinel 2 image covering an area of 585 sqkm in Forecariah region was analyzed and classified.

Land cover classes are adopted from Globeland30 types as water, wetland, agriculture, vegetation, bare-land and built-up areas, in the supervised classifications. While for the unsupervised classifications, the class number check was carried out. Train and test data were collected by using very high-resolution images under Google Earth close to Sentinel 2 image acquisition date. The thematic maps of the study area were obtained by using K, RF, SVM and DL4J. We run K as unsupervised method, under SNAP. For supervised methods, RF and SVM classifiers were carried out under EnMAP-Box in QGIS. Finally, DL4J was carried out under Weka. The classification results were obtained by using 6 clusters with 30 iterations and validated by the thumb rule for the unsupervised method K. On the other hand, for the 6 clusters, we applied 10-fold cross validation for the supervised methods. K, RF, SVM and DL4J delivered overall accuracies of 78%, 83%, 86% and 91% respectively. DL4J gave the best result which can be used for generating future LULC change maps in the area by using Sentinel 2 images.

## 4.2 Recommendations

Even-though DL4J gave the best outcome in terms of result among the studied classifiers, the disadvantage of the method is its high computational cost leading to longer processing times. It is necessary to keep searching until we find out some other DL algorithms that best fit the research objectives. This should be free of charge and less computational time demanded classifier. Monitoring the LULC features of developing regions by using Sentinel 2 images with strong open-source classifiers can be beneficial for public policy makers. Thus, a change map will be created and updated at certain intervals which will be utilized to keep tracking LULC and climate changes. These maps can be used within a decision-based system which then will improve the land administration capability of the authorities.

Further to this study, what we can suggest as points are:

- 1- Accentuate the use of DL methods in the RS projects,
- 2- Build strong decision-making systems via open-source algorithms,

- 3-** Make use of many different free of costs programs to monitor LULC features,
- 4-** Mostly use supervised methods in the classification of urban area and climate monitoring
- 5-** Mostly use open data sources for the change tracking in developing countries
- 6-** Find out some other DL algorithms that best fit sporadic urbanized zones, free of charge that demand less computational time,
- 7-** Further use change detection maps to facilitate both LA and climate monitoring process.



# References

1. James Anderson R, Ernest hardy E, John Roach T, Richard Witmer E. A land Use and Land Cover Classification System for Use with Remote Sensor Data. Geological Survey Professional Paper; 1983
2. Larsson G. Land management as public policy. University Press of America; 2010
3. Hofmann-Wellenhof B, Lichtenegger H, Wasle E. GNSS—global navigation satellite systems: GPS, GLONASS, Galileo, and more. Springer Science & Business Media; 2008.
4. Praveen kumar M, Jayarama Reddy Sreenivasula R. Analysis of Land Use Land Cover Changes using Remote Sensing Data and GIS at an Urban Area, Tirupati, India. Scientific World journal. April, 29th 2013.
5. van Maarseveen M, Martinez J, Flacke J. GIS in sustainable urban planning and management: a global perspective. Taylor & Francis; 2019.
6. Hatzopoulos JN. Topographic Mapping: Covering the Wider Field of Geospatial Information Science & Technology (GIS & T). Universal-Publishers; 2008.
7. Dale P, McLaughlin J. Land administration. Oxford University Press; 2000.
8. Zhu Z, Wulder MA, Roy DP, Woodcock CE, Hansen MC, Radeloff VC, *et al.* Benefits of the free and open Landsat data policy. Remote Sensing of Environment. 2019; 224: 382-5. doi.org/10.1016/j.rse.2019.02.016

9. Aschbacher J. ESA's earth observation strategy and Copernicus. Ed.: Onoda M and Young OR. Satellite earth observations and their impact on society and policy. Springer; 2017. 81-86.
10. Drusch M, Del Bello U, Carlier S, Colin O, Fernandez V, Gascon F, Hoersch B, Isola C, Laberinti P, Martimort P, Meygret A. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote sensing of Environment. 2012; 120: 25-36.
11. Regards OSS. Spot World Heritage (SWH) Data Site [Internet]. Toulouse; 2021 [erişim tarihi 22.06.2021]. <https://regards.cnes.fr/user/swh/modules/60>.
12. Marco C, Giovanni P, Carlo S, Luisa V. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. Computer Science. 2015.
13. Packt P. Introduction to Clustering and unsupervised learning. Data: Artificial Intelligence News feb, 23rd 2016.
14. Lemenkova P. Applying Isodata and K-means Algorithms of Unsupervised Classification Methods of ENVI for Airborne CASI Image Processing. December, Research Gate 2009.
15. Akinwumi A, Angel G, Helen C. African Economic Outlook: Entrepreneurship and Industrialization. New York: OECD; 2017.
16. Tony Y. Understanding random forest: *How the algorithm works and why is it so effective*. Data science June12th, 2019.
17. Onesmus M, Lalithnarayan C. Machine Learning: *Introduction to random forest in machine learning*. Section. December 11th, 2020.
18. Sunil R. Understanding Support Vector Machine algorithm from examples (along with codes). Analytics Vidhya. September 13<sup>th</sup>, 2017.

19. Yalcioglu S. Classical Examples of Supervised vs Unsupervised learning in Machine Learning. Springboard 2020.
20. Cliff M. Grids and Datums Republic of Guinea. Photogrammetric Engineering & Remote Sensing. 2002
21. United State Department of Agriculture. MAPASYST: Geospatial Technology. *What's the difference between a supervised and unsupervised image classification?* Available from: [http://mapasyst.extension.org/whats\\_the\\_difference\\_between-a\\_supervised\\_and\\_unsupervised\\_image\\_classification/](http://mapasyst.extension.org/whats_the_difference_between-a_supervised_and_unsupervised_image_classification/) [Accessed in August, 2019].
22. Traore, A, Watanabe T, Wentz E. Modeling Determinants of Urban Growth in Conakry, Guinea: A Spatial Logistic Approach. Urban Sci. 2017.
23. Amadou M, Merkel A. *Climate Conakry*. Available from: <https://en.climate data.org/africa/guinea/conakry/conakry-527/> [Accessed 19th February 2019].
24. Chen J, Liao A, Cao X, Chen J, Chen L, Chen X, He C, Han G, Peng S, Lu M, Zhang W, Tong X, Mills J. Remote Sensing mapping of global land cover. Beijing: Science Press; 2015; (103):7-27.
25. Sebastian VDL, Andeas R, Mathias H, Benjamin J. The EnMAP-Box— A Toolbox and Application Programming Interface for EnMAP. Research-gate. Sep 2015.
26. Steven C. Deep Learning for beginners: A comprehensive introduction of deep learning fundamentals, for beginners to understanding frameworks, neural networks, large datasets, and creative application with ease. SCRIBD: Data Science; November 6<sup>th</sup>, 2018.
27. Lan HW, Eibe F, Mark AH Christopher JP. Data Mining: Practical machine learning tools and techniques. Fourth Edition. Morgan Kaufmann: Elsever 2017.

# Curriculum Vitae

Name Surname : Yousseuf Camara  
E-mail (1) : y180228018@ogr.ikcu.edu.tr  
E-mail (2) : [yousseuf@balciogluinsaat.com.tr](mailto:yousseuf@balciogluinsaat.com.tr)  
E-mail (3) : youbenibra@gmail.com

## Education:

2014–2018 BSc., İzmir Kâtip Çelebi University, Dept. of Geomatics Eng.  
2018– MSc., İzmir Kâtip Çelebi University, Dept. of Geomatics Eng.

## Work Experience:

2019 – 2020 General Manager of Balcioglu Constructions A.Ş (Office of Guinea)

## Publications:

1. Land use/land cover classification of sentinel 2 image: Forecariah, Guinea, case study". 3rd International Students Science Congress (03-04/2019).

*link: <http://www.sciencecon.org>*

2. Architectural works and monuments of our world (10/03/2021)

*link: <http://www.udef.org.tr/tr/makaleler/yayinlarimiz/>*