

Urbanization Change Analysis based on SVM and RF Machine Learning Algorithms

Farhad Hassan¹, Tauqeer Safdar², Aman Ullah Khan⁴
Department of Computer Science & Engineering
Air University Multan Campus, Multan, Pakistan

Syed Muhammad Husnain Kazmi⁵
IAR Systems AB
Strandbodgatan 1, 75323 – Uppsala, Sweden

Ghulam Irtaza³
Department of Computer Science, University of Education
Multan Campus, Multan, Pakistan

Farah Murtaza⁶
School Education Department
Multan, Pakistan

Abstract—To maintain sustainability in the development, measured the yearly change rate of the land through Land Cover classified maps that hold the data which is surveyed as an influential factor for environment management and urbanization. This paper measured the change rate, which is helpful for the management of the city to define the new policy and implement the best one to maintain the natural resources. Machine Learning algorithms are utilized to produce the most acknowledged Land Cover maps using the GEE cloud-based reliable platform using the LANDSAT8 satellite imagery. For the classification used the Random Forest (RF) and Support Vector Machine (SVM) Algorithm. This investigation also found that the Support Vector Machine (SVM) classifier accomplished better over-all accuracy and Kappa coefficient as compared to the Random Forest (RF) classifier while the training sample for both is the same.

Keywords—Random Forest (RF); Support Vector Machine (SVM); GEE; classification; machine learning classifier; multi-temporal change analysis; urban change analysis; LANDSAT8; Kappa co-efficient

I. INTRODUCTION

Urbanization is a global phenomenon with various nations encountering different frequencies and precedents of rural-urban relocation. The whole world is changing into an urbanized center with a more significant part of the individuals moving to huge urban communities during the previous decades. Almost everyone is either way engaged with this procedure. Before mid-century, urban advancement was, for the most part, limited to developed nations yet has spread to developing nations now. Presently, practically all the developing nations on the planet are encountering urbanization. Advancing investigation suggests that the association between urbanization and advancement is not customized. Urbanization is ordinarily seen as solidly related to fiscal improvement, particularly in advanced countries. It is evaluated that urban territories make over 80% of the overall GDP. Enormous scale urbanization is changing the traditional structure prompting a significant upheaval in our general public. The development of urbanization in Pakistan mirrors the desires and ambitions of billions of new urbanites. It is a repetitive procedure a nation encounters as it advances from an undomesticated to modern culture. People migrate from

monetarily restrained territories to a spot where fitter possibilities are advertised [1].

Pakistan has an average level of urbanization inside the Asia-Pacific area, dependent on both the development of urbanization and urban evolution. Pakistan has an enormous shelter shortage of around ten million units and increasing. Urban masses' improvement in the country has not been facilitated by advancement in shelter units or unbiased access to land, causing shelter to lack and the preferment of ghettos. Given Pakistan's shortcoming to earthquakes and other natural hazards, the lawmaking organization should set up a practical framework for developing guidelines and their consistency to avoid the implied adverse effects of vertical shelter schemes. The modern approach to manage urban shelter has faced multiple challenges. Although, among the South Asian nations, Pakistan has the most raised portion of the population living in the urban zone. The urbanized percentage of Pakistan's population grew 4.1 percent in the five years between 2005 and 2010 and is expected to grow more in the near future years [2]. As per the UN, a fraction of the populace will be residing in urban areas by the year 2030. Altogether, urban networks in Pakistan contribute fifty-five percent of the GDP. Furthermore, Pakistan generates 95% of its government salary from 10 substantial urban zones [2].

A Multan district is in the southern area of the Punjab province, Pakistan. Multan is also famous with the name City of Saints and is found on the banks of the River of Chenab at 30.1575° N latitude and 71.5249° E longitude. Total covering area around 3,721. Sq.km and it is the seventh-largest city in Pakistan. Urbanization is the transformation of any land into urban land and globally, it is observing that the trend of urbanization is increasing rapidly. To measure this change needs the upper layer view of the Earth's surface like satellite images. Numerous sensors are catching tremendous volumes of high-resolution satellite images containing diverse information that is updated diurnally. Remote sensing is the process of receiving the information associated with the Earth by performing the scanning process with the assistance of satellites. Remote sensing is performed to achieve the precise and clear image of the Earth, and these high-resolution remote sensing images are free of any cost, and for the researcher and

analyst, it is an incredible possibility to achieve better analysis results in a specific region of interest (Lopez, R. D., & Frohn, R. C. (2017)). If these images utilized in an appropriate manner, then it gives a tremendous profit and investigates the surface of the Earth [1]. As per a near perspective, remote sensing imagery contrasts from the standard imagery, the significant explanation behind these matters is additional spectral information that the satellite imagery contains and which is not noticeable for the unaided eye [3]. On the high-resolution satellite images, Land Use Land Cover (LULC) change examination is one of the major because from this investigation report obtains the yearly trend of urbanization change that helps the policymaker of any establishment makes the correct choice and makers additionally understand the impacts of these progressions on people and their environments. Land Use and Land Cover (LULC) investigation has an exceptional influence on the future of the land, and this examination depends on the classification of the particular feature's class [4]. Land Cover Change Analysis screens the change either in loss or gain of any land type over the time range and measures the status of progress by the analysis for the scope of course of events. Yu et al. (2016) Land use have frequently affected land cover. The surface of the land has been changed in the part of appearance when seen on two distinctive time-series [5]. Furthermore, Huang et al. (2017) mentioned that Google Earth Engine (GEE) is the reliable and powerful platform created by Google, Carnegie Mellon University and the US Geological Survey, uninhibitedly accessible; a propelled distributed computing conditions for remote sensing data processing and geospatial analysis and no need of any fast CPUs or GPUs for high computational work aside from fast internet connection [6].

The paper is organized into sections like problem statement which describes the land loss due to urbanization. In the next section, a literature survey is done where the machine learning approaches are mentioned; later in that section, satellite image collection and analysis processes are mentioned by highlighting the work of different researchers. In the later section of methodology, the data collection methods and SVM and RF algorithms are used to obtain the results of the study which describes the transformation of land in 2017 and 2018 and its impacts. In the last section the paper is concluded by mentioned the land transformation changes.

II. RESEARCH PROBLEM

The continuous urbanization may result in several concerns for the supervision of the city. In order to increase the sustainable progress of the city, there is a demand for measuring Land Cover Change analysis. From this investigation, obtain the influence of water-bodies and vegetation land on the advancement of urbanization with the help of the SVM and RF machine learning algorithms during the multi-temporal range of 2017 to 2018. The research is conducted for the lack of land cover change analysis in order to measure the change of urbanization. The motivation behind the research is to measure the changes in water bodies and vegetation using the classifier (SVM and RF) approach. Through the latest cloud-based powerful and the most reliable environment named, GEE and designing the new policies for the accommodation of the development.

III. LITERATURE REVIEW

Agriculture is a practice of producing and harvesting of the crops in a methodical manner. Because of the ever-increasing demand for the food, there is a need to improve the yield of the crop and avoid the loss of crop in every possible manner. In order to do so, the scientific community is using the optimal resources for the cultivation of crops, and remote sensing is used for this purpose because of its enormous advantages and features. One of the most vital applications of remote sensing is the classification of the crops by differentiating between the varieties of the crop.

Varma et al. (2017) Satellite imaging are used for the viable investigation of the temporal changes that can affect the yield of the crop in the specific areas. By using satellite imaging methods, the growth of crops from sowing to the harvesting level can be monitored in a very efficient way. Geo-references and ortho-rectified satellite images are used to identify the land loss problems and the affected areas in a different region of the world, not just that the seasonal changes and crop variation can also be monitored by using these techniques. Moreover, on the information based on the activities like deciding the type of crop and its acreage, growth determination of different stages of the crop, and delineating the extent of the crop can be planned in advance. These methods are then used for the decision-making policies to increase the yield of the crop and avoid land loss.

Senthilnath et al. (2016) Another satellite image collection method called multi-spectral satellite imaging can aid the identification and classification of the crops because they take chances in the reflectance as a function for the specific crop type. The classification of the crop is used in the auditing of the land use, and it also facilitates the soil and water study of the particular area. Because of the variability in the cultivation of the crop in a geographical area the process of the classification is a very challenging factor. Moreover, the classification of the crop is done on the basis of spatial and spectral bands or also by combining both methods for the purpose of classification. The clustering method is used for the grouping of the explicit datasets in such fashion that the data point towards the same group is virtually analogous. The purpose of clustering is that it reduces the intra-cluster distances, and it maximizes the inter-cluster distance. The data extracted from the points are in the form of optimal cluster centres, and the clustering is performed on the satellite data in several ways, for example, the clustering methods are focused on the partitioning using the spatial patterns. Partition clustering is completed by dividing the data with the fixed number of clusters (priori) by using the similarity indexes. K-means is known as one of the most commonly used clustering methods.

Machine learning is one of the most unswerving classification methods. In machine learning, the classifier works to determine the category to which a new research sample belongs based on the instruction data of the marked kind. The machine learning technique makes its technique more suitable in classification, especially in the field of remote sensing images because it is impossible to have a detailed knowledge area of study. Satellite images broadly utilized for

land cover change investigation and more often than not the manager and policy designer need a reliable technique to get to the result of a land cover change by reviewing and recognizing changes of land types Karpatne et al. (2016).

The research yield portrays the diverse land includes that is available increases and lost on the surface of the Earth. GEE gives the most precise and dependable research data with a financially savvy parameter and takes lesser time contrasted with different gadgets that can be utilized for land cover change analysis. Mueller et al. (2017) reported that Landsat Satellite imagery is profitable for the various plans, for example, to measure the change of climate and disasters related to the environment and management of natural resources. Land cover change analysis is the under of natural resources either created by any machine learning classifier, threshold method or the hybrid approach. Landsat data are most commonly used in natural phenomena-oriented applications.

Borra et al. (2019) reported that the procedure of satellite image classification encompasses the gatherings of image's pixel esteem into the section of feature assortment, and for the satellite image classification, various strategies are accessible [7] to create the map. These maps can acquire various highlights or qualities to do the ideal work. A few different strategies incorporate the supervised and unsupervised plan to acquire the estimation of the precision of remote sensing information which is an essential requisite in classification.

Support Vector Machine Classifier is among one of the many mainstreams and the unanimously acknowledged classifiers in the domain of remote sensing, the essential purpose for this is profoundly precise classification results. It is the binary classifier and dependent on the concept that the training samples which are nearer concurrence to the boundaries of a class will differentiate the class superior to other samples of training [8]. So, it is evident that SVM classifier focuses on finding the optimal hyper-plane that separates the samples of training input into many classes and samples of training data is close to the boundaries of the class and at the lesser distance to hyper-plan are taken as support vectors, which is to be used for actual training [12]. Maxwell, A. E. (2018) the selection of kernels plays a significant role in the classification's results. RBF is also the type of kernel in which has a user-defined parameter that controls the impact of a sample of training on the boundary of decision and the value of the user-defined parameter is higher, then there is the chance of over-fitting. So that is why it is necessary to take the right balance for value [9]. Tehrany et al. (2015) conducted a study where the researchers implemented the Support Vector Machine (SVM) algorithms in different fields of study like flood susceptibility assessment and landslide susceptibility investigation. Genetic algorithms are acknowledged as the most advanced and pervasive developed heuristic search models in the field of artificial intelligence, and it has its application in the urban planning, ecological studies, climate modelling, and remote sensing. Kruber et al. (2019) also evaluated the performance of machine learning algorithms called RF and SVM and compared the results with each other and mentioned that RF and SVM are called a mapping

algorithm for the groundwater outbreaks, and these algorithms can provide very effective results.

IV. AREA OF STUDY

A Multan district is in the southern area of the Punjab province, Pakistan. Multan is also eminent with the name City of Saints and is found on the banks of the River of Chenab at 30.1575° N latitude and 71.5249° E longitude. Total covering the area around 3,721. Sq.km and it is the seventh-largest city in Pakistan. The variety of feature classes present in Multan makes it a good candidate for the study, and the selected area of study for this research is 756.52368 sq. kilometers. The coordinate's values of area of study are [71.34246, 30.29730], [71.34246, 30.08839], [71.68166, 0.088396] and [71.68166, 30.29730]. Fig. 1 indicates the location of the Multan District under the boundaries of Pakistan and also shows the selected area study of area for the research.

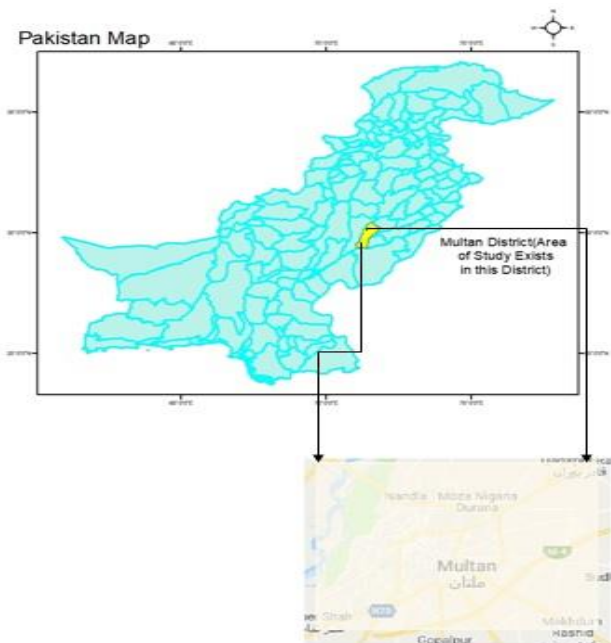


Fig 1. Area of Study.

V. DATA PREPROCESSING

The satellite imagery data use in the research is Landsat 8, Collection 1, Tier 1 and ID of this dataset is LANDSAT/LC08/C01/T1. The quality of this dataset is high, and it is considered a more suitable multi-temporal analysis. The total number of bands of this dataset is 12 and for this research just selected just 6 bands B2, B3, B4, B5, B6, and B7. The area bound was set according to the above-mentioned coordinate's value under the district of Multan. For the year 2017, the collection of images is 43, and for 2018 the numbers of an image are 45, and these 88 images are filtered as per area of study. After this select only two images from them one for 2017 and one for 2018. For this reduced selection, convert them into TOA reflectance, simple cloud score to be calculated and in the last apply the median on the least cloudy as shown in Fig. 2.

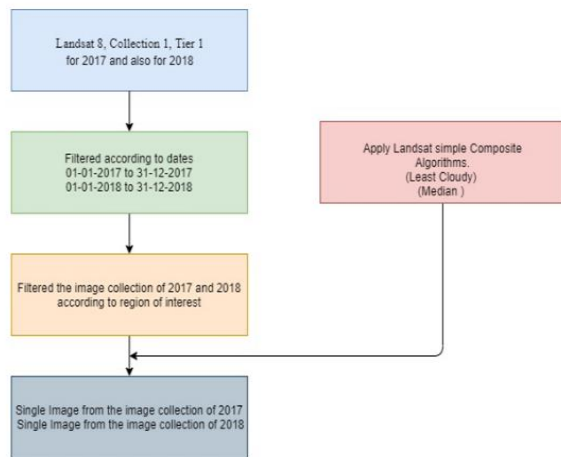


Fig 2. Data Preprocessing.

VI. METHODOLOGY

To attain the urbanization of yearly drift the adopted methodology is shown in Fig. 3. From the LANDSAT8 take the image assortment and in 2017 assortment of satellite images are 43 however in 2018 the assortment which comprises of 45 images as indicated by the filtration of the study zone limit, then apply the composite algorithm which is exceptionally valuable to make an assortment of images into a solitary image. The composite calculation forms the median from the assortment of Landsat images and the exclusive image from the assortment ought to be the sans cloud image. The composite algorithm pertains to both image assortments of 2017 and 2018.

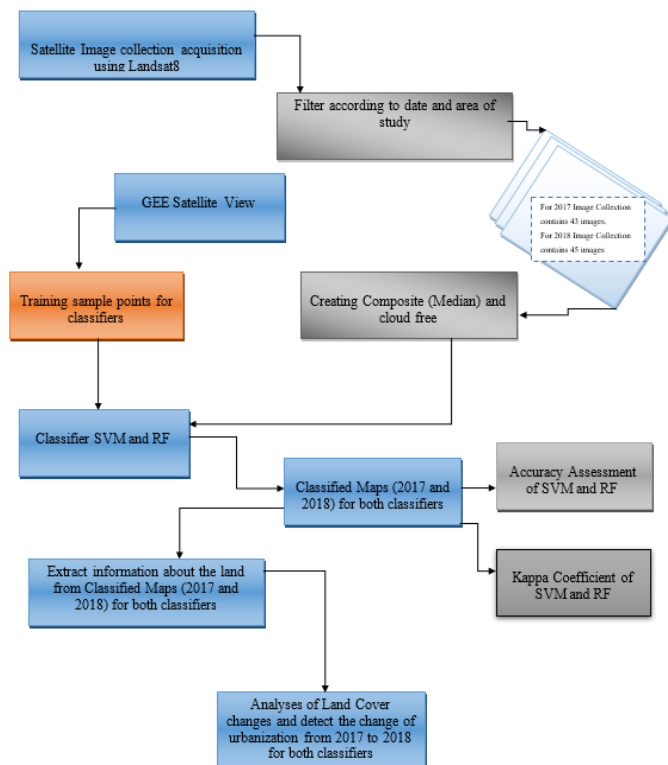


Fig 3. Research Methodology.

For the determination of the representations of training points, utilizing the geometry alternative such as polygon and point of the GEE frame for the classification and uses two unique classifiers SVM and RF to compose the classified maps as shown in the Fig. 3. Estimate the precision and kappa coefficient of both SVM and RF classifiers and additionally anticipates the varieties of land in square kilometer for 2017 and 2018. In the final prognosis, the variation in urbanization land from different types of land over time, operating the two classifiers.

VII. RESULTS AND ANALYSIS

The characterized images for years 2017 and 2018 are created by applying machine learning classifiers (SVM and RF) for the apprehension of the variation in urbanization. We use SVM and random forest classifier on a comparable training data index. We analyze that the consequences produced by the SVM classifier are more exact than the random forest classifier. SVM classifier produces the effects by retaining the pixel's fallacious quality with low miss arrangement of urbanization.

The projected results for the year 2017 and 2018 utilize the SVM classifier for the classification with the subtleties of three sorts of land water, vegetation and urban under the selected area of study. Fig. 4 and 5 are representing the classified results using SVM classifier.

With the utilization of the RF classifier, created the detailed map for the year 2017 and 2018 are appeared in Fig. 6 and 7, these show the outcomes for 2017 and 2018. Both resultant images include three distinct kinds of land water bodies, vegetation and urban. Fig. 6 and 7 are representing the classified results using RF classifier.

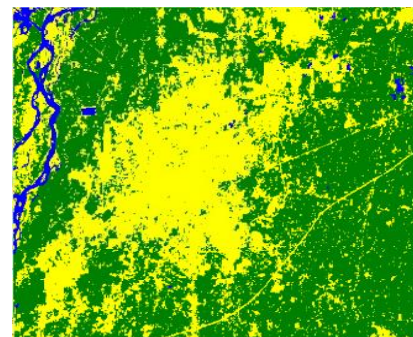


Fig 4. Classified Results for 2017 using SVM.

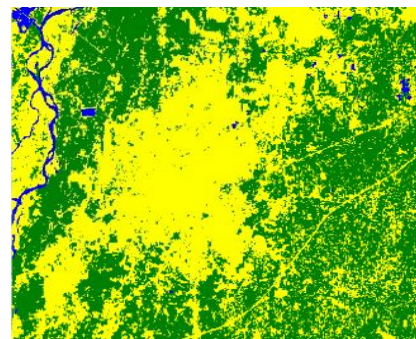


Fig 5. Classified Results for 2018 using SVM.

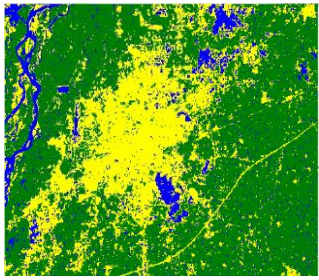


Fig 6. Classified Results for 2017 using RF.

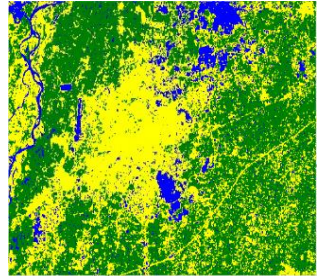


Fig 7. Classified Results for 2018 using RF.

a) Type of Land for 2017

Extracted each type of land from the classified results under the study area using the machine learning (SVM and RF) classifiers. The water area for the year 2017, which is obtained from the SVM and RF classifier is shown in Fig. 8 and 9, respectively. Blue color indicates the existence of water under the area of study while the black color indicates any other land type area.

The vegetation or greenish area as presented in the Fig. 10 and 11 identified using the Green color, indicates the existence of vegetation while black color indicates the non-vegetation area under the area of study.

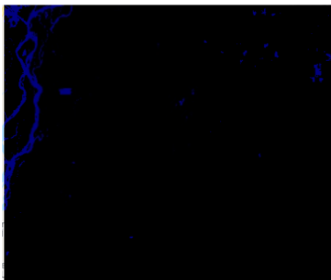


Fig 8. Water 2017 based on SVM.

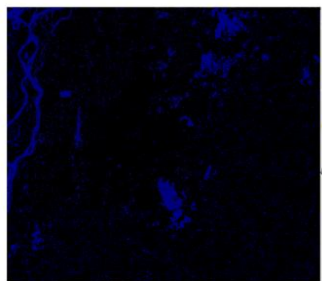


Fig 9. Water 2017 based on RF.

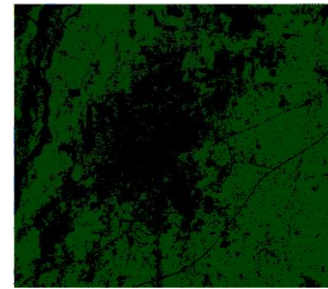


Fig 10. Vegetation Area in 2017 based on SVM.

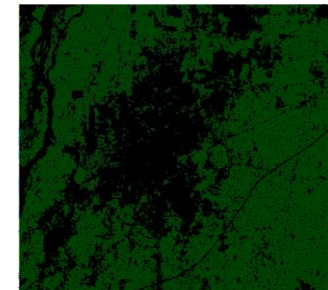


Fig 11. Vegetation Area in 2017 based on RF.

The urban area for the year 2017 which is obtained from the SVM classifier results is shown in Fig. 12 and urban area for the year 2017 which is obtained from the Random forest classifier is shown in Fig. 13. In both, Yellow color indicating the area of urban land presence while the black color indicates any other land type area.

b) Type of Land for 2018

For 2018 extract each type of land from the classified results under the study area using the machine learning (SVM and RF) classifiers. The water area for the year 2018 which is obtained from the SVM and RF classifier is shown in Fig. 14 and 15, respectively. Blue color indicates the existence of water bodies under the area of study, while the black color indicates any other land type area.

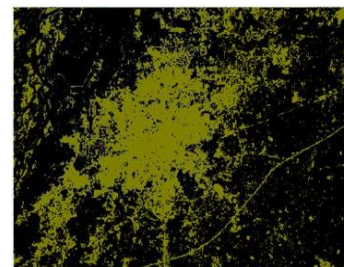


Fig 12. Urban Area in 2017 based on SVM.

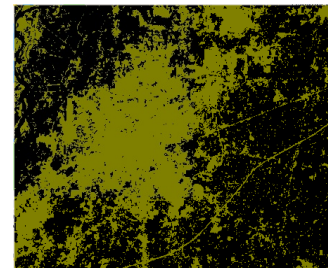


Fig 13. Urban Area in 2017 based on RF.

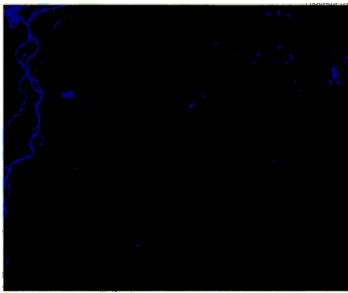


Fig 14. Water 2018 based on SVM.

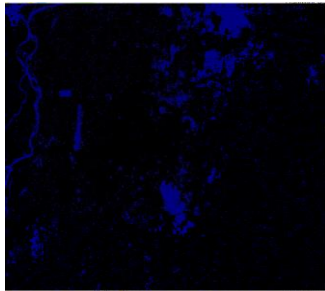


Fig 15. Water 2018 based on RF.

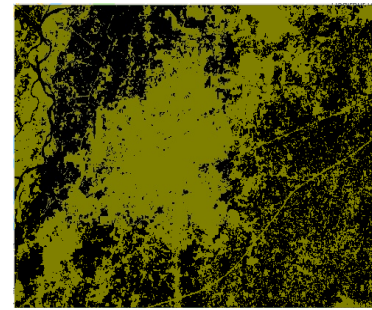


Fig 18. Urban Area in 2018 based on SVM.

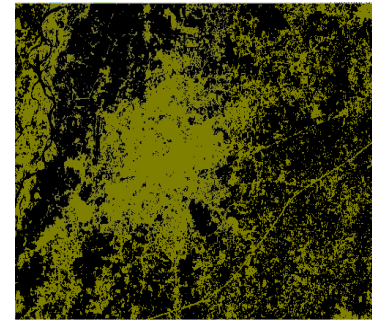


Fig 19. Urban Area in 2018 based on RF.

The urban area for the year 2018 which is obtained from the SVM classifier results are shown in Fig. 16 and urban area for the year 2018 which is obtained from the Random forest classifier is shown in Fig. 17. In both, Yellow color indicating the area of urban land presence while the black color indicates any other land type area.

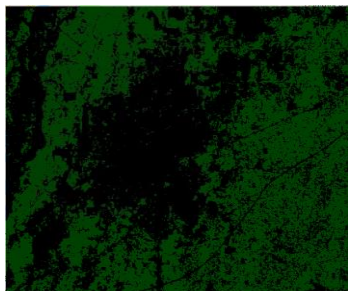


Fig 16. Vegetation Area in 2018 based on SVM.

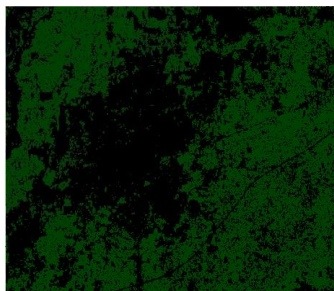


Fig 17. Vegetation Area in 2018 based on RF.

The urban area for the year 2018 which is obtained from the SVM classifier results are shown in Fig. 18 and urban area for the year 2018 which is obtained from the Random forest classifier is shown in Fig. 19.

1) *Measure Water Bodies Losses*: To calculate the water bodies' losses between the multi-temporal ranges from 2017 to 2018 firstly obtain the values of water bodies of both years separately by using both the classifier's results. In 2017, the area of water bodies according to the classified result of SVM is shown in Fig. 8 under the area of study while for the year 2018 is shown in Fig. 12. From both, extract the area of losses in waterbodies and it is shown in Fig. 20. While the loss area of water bodies based on RF classifier's results over the time series of 2017 to 2018 is shown in Fig. 21.

The area covered by water bodies that are obtained using the SVM classifier for the year 2017 is 17.733197689859043 sq. kilometer and in 2018 it is 10.974930358946 sq. kilometer under the area of study which is 756.5236827301217 sq. kilometer. The calculated difference between both extracted land type shows that the water bodies' area is decreased by 6.75826733091299 sq kilometers from the classified results of SVM and RF as shown in Fig. 20 and Fig. 21. Using RF classifier, we inspected that in 2017 water-bodies area is 71.40641692305495 sq. kilometers, in 2018 it is decreased by 8.188619262852404 sq. Kilometers and the remaining water area is 63.2177976602023 sq. Kilometers.

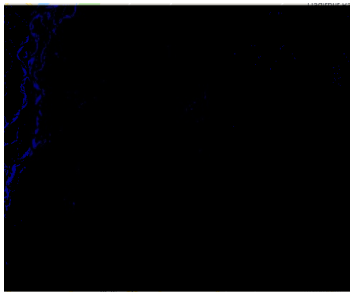


Fig 20. Waterbodies Losses Area 2017 to 2018 based on SVM.

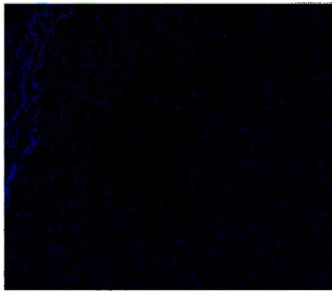


Fig 21. Water Bodies Losses Area 2017 to 2018 based on RF.

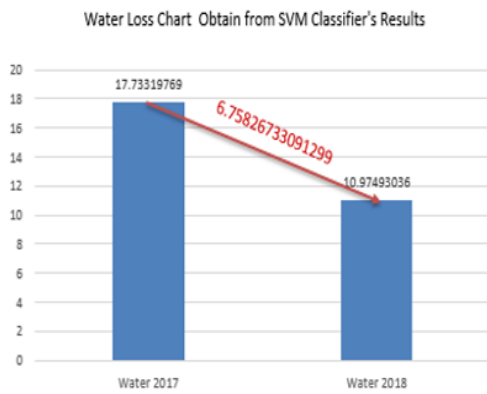


Fig 22. Water Loss Chart based on SVM.

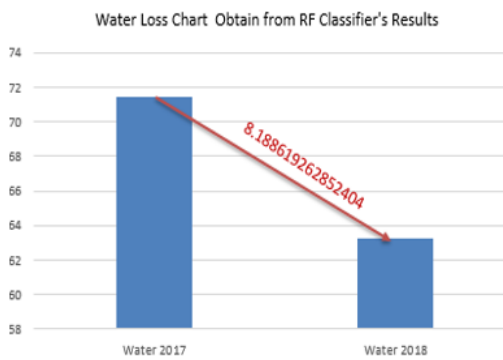


Fig 23. Water Loss Chart based on RF.

2) *Measure Vegetation Losses*: Extract the vegetation area from the classified resultant output of 2017 and 2018 using the SVM classifier and Random Forest, and then measure the difference from both years results to obtain the loss area of vegetation from 2017 to 2018 as shown the Fig. 22 and 23 based on SVM and RF.

The area covered by vegetation is obtained using the SVM classifier for the year 2017 is 435.1198872814169sq. Kilometers and for the year 2018, it remains 355.15992729514426 sq. kilometers under the area of study which are 756.5236827301217 sq. kilometers. The vegetation area decreased by 79.95995998627411sq. Kilometer, as shown in Fig. 24 and 25. Using RF classifier, in the 2017 vegetation area is 451.10853885372273 sq. kilometers but in 2018 it remains 362.966601021753 sq. kilometers. The loss of vegetation area from 2017 to 2018 is 88.14193783196974 sq. kilometer as shown in Fig. 26 and 27.

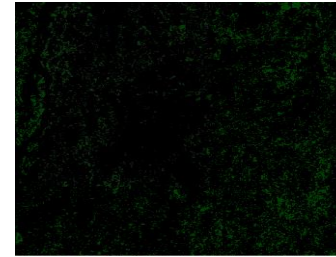


Fig 24. Vegetation Loss 2017 to 2018 based on SVM.

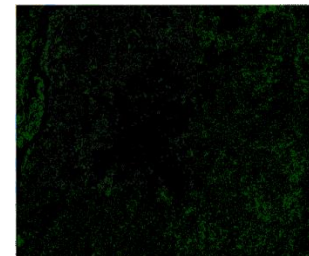


Fig 25. Vegetation Loss 2017 to 2018 based on RF.

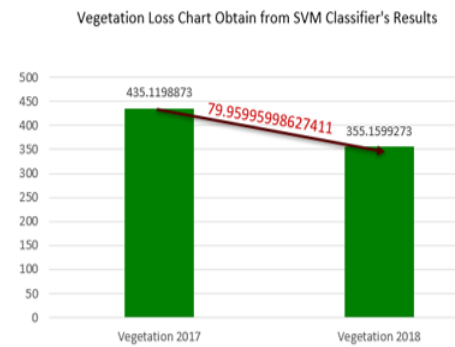


Fig 26. Vegetation Loss Chart based on SVM.

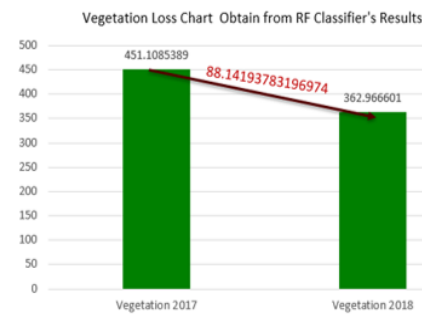


Fig 27. Vegetation Loss Chart based on RF.

3) *Urban Change*: From the classified results of SVM and RF measure the change in urban land over the multi-temporal range of 2017 to 2018. Fig. 28 shows the change of urban land-based on SVM, and Fig. 98 shows the changing area of land but for RF classifier.

From the classified resultants of SVM for the year 2017, the urban land area was 303.67059775885275 sq. Kilometers and in 2018 the area of urban land reaches 390.3888250760368 sq. kilometers. The urban land area increased by 86.71822731718704 sq. Kilometer using SVM, as shown in Fig. 28. Classified resultants of RF for the year 2017, the urban land area was 234.0087269533519 sq. Kilometers and in 2018 the area of urban land reaches 330.339284048179 sq. kilometers as shown in Fig. 29. The urban land area increased by 96.33055709482205 sq. Kilometers using RF and using SVM as shown in Fig. 30 and 31.

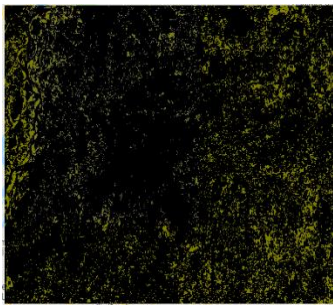


Fig 28. Urban Growth SVM.

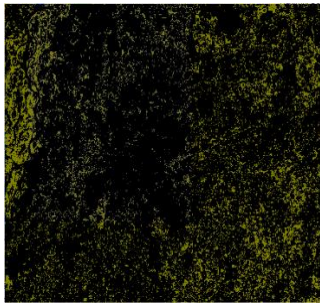


Fig 29. Urban Growth RF.

Urban Land Gain Chart Obtain from SVM Classifier's Results

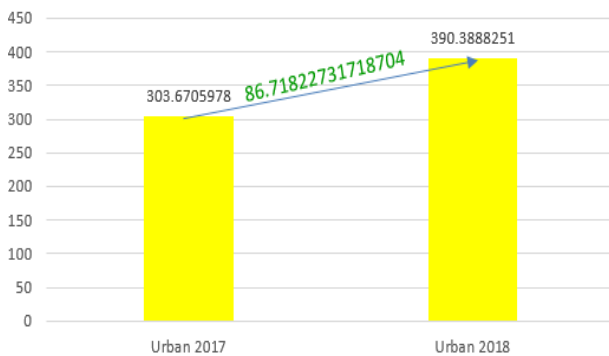


Fig 30. Urban Growth Chart based on SVM.

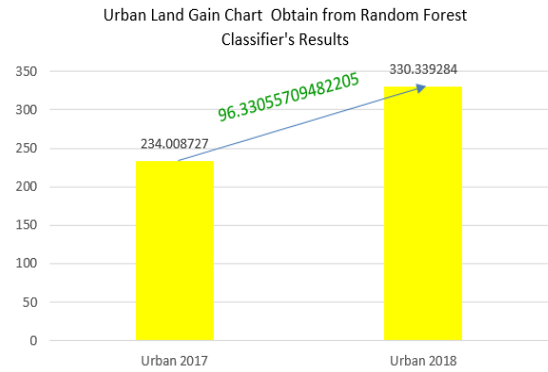


Fig 31. Urban Growth Chart based on RF.

a) Vegetation Land into an Urban land

From the classified results of both algorithms extract the vegetation land and obtained the transformation area of vegetation land into urban land. As shown in Fig. 32 for the SVM algorithm and Fig. 33 for the RF algorithm.

b) Water Bodies into an Urban Land

From the classified results of both algorithms extract the water-bodies and obtained the transformation area of water-bodies into urban land. As shown in Fig. 34 for the SVM algorithm and Fig. 35 for the RF algorithm.

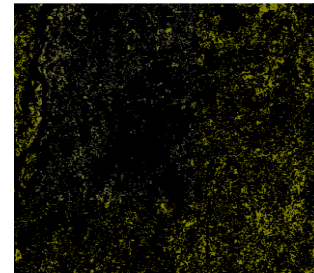


Fig 32. Urban in 2018 from Vegetation in 2017 based on SVM.

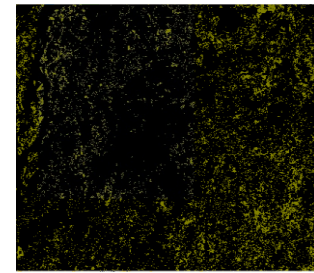


Fig 33. Urban in 2018 from Vegetation in 2017 based on RF.

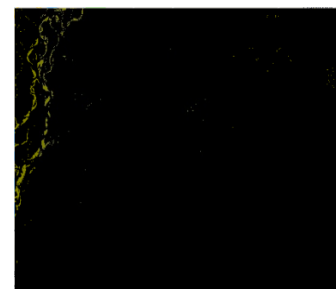


Fig 34. Urban in 2018 from Water in 2017 based on SVM.

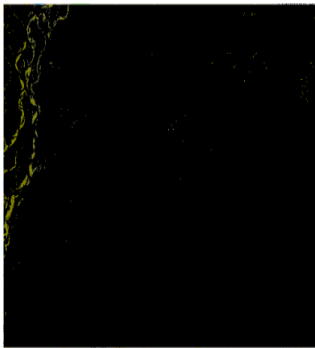


Fig 35. Urban in 2018 from Water in 2017 based on RF.

c) Total Change in Urban Land

The Total area of change from the vegetation loss and water bodies of 2017 in 2018 is extracted using the SVM, as shown in Fig. 36. While Fig. 37 indicates the total change based on RF.

From the classified resultants of SVM for the year 2017, the urban land territory was 303.67059775885275 sq. Kilometers and in 2018, the zone of urban land reaches 390.3888250760368 sq. Kilometers. The urban land territory extended by 86.71822731718704 sq. Kilometer utilizing SVM as shown in Fig. 38. Classified resultants of RF for the year 2017, the urban land zone was 234.0087269533519 sq. Kilometers and in 2018 the territory of urban land reaches 330.339284048179 sq. Kilometers. The urban land territory extended by 96.33055709482205 sq. Kilometers utilizing RF as shown in Fig. 39.

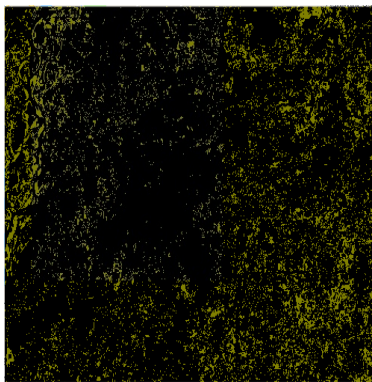


Fig 36. Total Urban Change based on SVM.

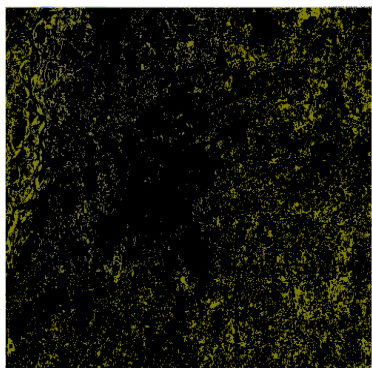


Fig 37. Total Urban Change based on RF.

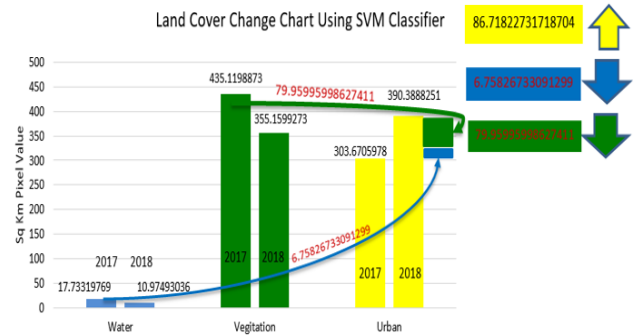


Fig 38. Land Transformation based on SVM Classifier.

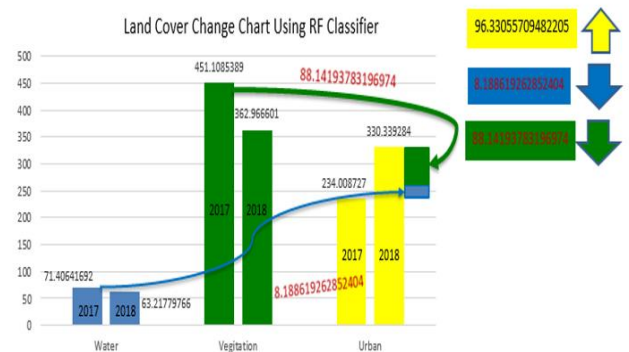


Fig 39. Land Transformation based on RF Classifier.

VIII. ACCURACY AND KAPPA COEFFICIENT ANALYSIS

1) Accuracy Assessment: For the assessment of accuracy uses the method of the Confusion matrix (error matrix). This method is the best approach for expressing accuracy of classification in supervised learning, and the error matrix (confusion matrix) provides the information of accuracy for every class in the classification [11].

$$\text{Classifier's accuracy} = (\text{TCP}/\text{TNP}) * 100$$

By using the above-mentioned equation, obtain the accuracy of the classifier and value of TCP and TNP can be calculated by a confusion matrix.

a) SVM Classifier Accuracy

To measure the accuracy of the model, use the values of the Confusion matrix (error matrix) as presented in Table I.

TCP is the sum of the diagonal values of the confusion matrix of the SVM Classifier.

$$\text{So, TCP} = 8+129+134=271$$

TNP is the sum of all values of the accuracy matrix.

$$\text{So, TNP} = 8+4+17+0+129+6+0+0+134=298$$

Put the calculated values of TCP and TNP into the above-mentioned equation.

TABLE I. CONFUSION MATRIX FOR SVM CLASSIFIER

	Water	Vegetation	Urban
Water	8	4	17
Vegetation	0	129	6
Urban	0	0	134

SVM Classifier's accuracy = $(271/298) * 100$
= **90.9395973154**

b) RF Classifier Accuracy

To measure the accuracy of the model, use the values of the Confusion matrix (error matrix). As given below here

Confusion Matrix for RF classifier

So, TCP = $3+38+35=76$

TNP is the sum of all values of the accuracy matrix.

So, TNP = $3+2+5+1+38+4+1+0+35=89$

Put values of TCP and TNP into the mentioned equation.

RF Classifier's Accuracy = $(76/89) * 100$
= **85.3932584269**

2) *Kappa Coefficient*: Kappa coefficient is used to control only those instances that may have been correctly classified by chance [10]. The value of this coefficient is between -1 to +1. The negative value of Kappa coefficient means low accuracy with between the classified image and the reference image, if the resultant value is 0 then it means no correlation at all between the classified image and the reference image and if the value is the higher then it indicates the higher accuracy of the classification [13].

Kappa Coefficient = $((OLA) - (ELA)) / ((TNP) - (ELA))$

The OLA means the Observed Level of Agreement which is basically the diagonal's values sum of the confusion matrix and ELA indicates the Expected level of agreement which is calculated by multiplying the total row values with total column values and the resultant value further divides by TNP repeats this till ends. After that sum up all values. TNP is the Total Number of Pixels of the confusion matrix. The confusion matrix for SVM classifier using the Kappa Coefficient value is calculated and presented in Table II.

a) Kappa Coefficient for SVM Classifier

OLA = $8+129+134 = 271$

ELA = row total * col total / overall total

Use the above-mentioned equation for each feature and sum all the obtained values

$8*29/298 = 0.7785234899328859$

$135*133/298 = 60.25167785234899$

$134*157/298 = 70.59731543624162$

So,

The sum of all these = 131.6275167785235 and

Expected Accuracy = 131.6275167785235

Put these calculated values in the Kappa Coefficient Equation,

Kappa Coefficient = $(271-131.6275167785235) / (298-131.6275167785235)$

Kappa Coefficient = 0.8377135480747897

The value of kappa Coefficient is in the range of 0.81–0.99, which shows almost perfect agreement.

b) Kappa Coefficient for RF Classifier

OLA = $3+38+35=76$

ELA = row total * col total / overall total

Use the above-mentioned equation for each feature and sum all the obtained values

$3*10/89 = 0.3370786516853933$

$40*43/89 = 19.32584269662921$

$44*36/89 = 17.79775280898876$

So,

The sum of all these = 37.46067415730336 .

Expected Accuracy = 37.46067415730336

Put these calculated values in the Kappa Coefficient Equation,

Kappa Coefficient = $(76-37.46067415730336) / (89-37.46067415730336)$

Kappa Coefficient = 0.7477654240244169

The value of kappa Coefficient is in the range of 0.61–0.80 that shows substantial agreement.

SVM classifier accomplished 90.9% comprehensive classification precision with a kappa estimation of 0.83, while the Random Forest (RF) classifier acquired 85.3% in general classification precision with a kappa estimation of 0.74. RF. SVM classified maps are adjoining to the ground facts when contrasted with RF generated maps. The RF map contained miss-classified water pixel information. We expect to concentrate on urbanization development from 2017 to 2018. Using SVM, the yearly change is 86.71822731718704 sq. kilometers and using RF it is 96.33055709482205 sq. kilometer.

TABLE II. CONFUSION MATRIX FOR SVM CLASSIFIER BASED KAPPA COEFFICIENT

	Water	Vegetation	Urban
Water	3	2	5
Vegetation	1	38	4
Urban	1	0	35

IX. CONCLUSION

The urbanization has affected the plans of settlement and use of the land around the sub-urban, increasing the population and areas concurrent. And by this mean, there have occurred the changes in the weather pattern raising the global warming and energy crisis. So, this study is conducted for the purpose to have a view about the land loss and how it is contributing to the energy crisis and greenhouse damaging. Analyzing the multi-temporal land cover change is very helpful for higher management and environmental departments to maintain sustainability in the development of urbanization. Realize the current change rate situation, then made the decisions and policy according to it. Also, for forecasting possible future changes with respect to growth to urban land/losses of water bodies and vegetation to avoid from any natural environmental problems that come in the near future. Uses the same size of training sample points and polygons for both classifiers and according to overall-accuracy and kappa coefficient comparison, observed that the SVM performed better as compared to RF, though both are superlative in performance when training sample data are smaller in size. Along with this, SVM classified maps are almost to the ground facts as compared to RF generated maps. The RF map contained miss-classified water pixel data. We utilize SVM and RF classifiers to assess the variation in urbanization. SVM and RF utilizing the prevailing platform GEE under an associated size of training sample points and polygons. Conferring to general precision and kappa-coefficient analysis, we observed that the SVM functioned better when contrasted with RF, while both are best in execution when training sample data are diminutive in size. This yearly change trend will be helpful for higher management and environmental departments to maintain sustainability in the development of urbanization by realizing the current situation, made better decisions and predicted future changes with respect to growth/loss land patterns including natural or unnatural environmental problems. The classification results of SVM algorithm are almost to the ground facts as compared to RF classification results and the majoris that the random forest classification results contained miss-classified water pixel data.

X. FUTURE DIRECTIONS

- 1) For the future work we can do the comparison of the results we obtained from supervised classifiers with unsupervised classifier results.
- 2) Digital image classification of multi-spectral imagery for prominent surface features, particularly with reference to population growth, within the precincts of Multan district can be done in the future.
- 3) Using Satellite imagery measure the impact of water-bodies growth/losses on vegetation for the same area of study.

REFERENCES

- [1] Liss, B., M.D. Howland, and T.E.J.J.o.A.S.R. Levy, Testing Google Earth Engine for the automatic identification and vectorization of archaeological features: A case study from Faynan, Jordan. 2017. 15: p. 299-304.
- [2] Blank, J., Clary, C., & Nichiporuk, B. (2014). Drivers of long-term insecurity and instability in Pakistan: Urbanization. Rand Corporation.
- [3] Ishii, T., et al. Detection by classification of buildings in multispectral satellite imagery. in 2016 23rd International Conference on Pattern Recognition (ICPR). 2016. IEEE.
- [4] Lam, N.S.-N., Methodologies for mapping land cover/land use and its change, in Advances in land remote sensing. 2008, Springer. p. 341-367.
- [5] Mountrakis, G., et al., Support vector machines in remote sensing: A review. 2011. 66(3): p. 247-259.
- [6] Thunig, H., et al. Land use/land cover classification for applied urban planning-the challenge of automation. in 2011 Joint Urban Remote Sensing Event. 2011. IEEE.
- [7] Borra, S., R. Thanki, and N. Dey, Satellite Image Clustering, in Satellite Image Analysis: Clustering and Classification. 2019, Springer. p. 31-52
- [8] Mantero, P., et al., Partially supervised classification of remote sensing images through SVM-based probability density estimation. 2005. 43(3): p. 559-570.
- [9] Maxwell, A.E., T.A. Warner, and F.J.J.J.o.R.S. Fang, Implementation of machine-learning classification in remote sensing: An applied review. 2018. 39(9): p. 2784-2817.
- [10] Karpatne, A., Jiang, Z., Vatsavai, R. R., Shekhar, S., & Kumar, V. (2016). Monitoring land-cover changes: A machine-learning perspective. IEEE Geoscience and Remote Sensing Magazine, 4(2), 8-21.
- [11] Thanh Noi, P. and M.J.S. Kappas, Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. 2018. 18(1): p. 18.
- [12] Yang, J., et al. Attribute weighted Naive Bayes for remote sensing image classification based on the cuckoo search algorithm. in the 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC). 2017. IEEE
- [13] Rwanga, S. S., & Ndambuki, J. M. (2017). Accuracy assessment of land use/land cover classification using remote sensing and GIS. International Journal of Geosciences, 8(04), 611.