

PERFORMANCE ASSESSMENT OF CLASSIFICATION ALGORITHMS FOR LANDUSE / LANDCOVER CHANGE USING SENTINEL 2 DATA - A CASE STUDY OF TIRUPPUR

B. Prawin ^a, P. Masilamani ^{a*}, S. Abdul Rahaman ^b

^a Department of Geography, Bharathidasan University, Tiruchirappalli- 620 024.- prawin24geo@gmail.com ,
masilamani.ap@gmail.com

^b Department of Geography, Bangalore University, Bengaluru- 560056-abdulatgeo@gmail.com

Commission VI, WG VI/4

KEY WORDS: Sentinel-2, LULC, K-means, IsoData, support vector machines, and maximum likelihood.

ABSTRACT:

In the history of mankind, one of the vibrant geographical phenomena is urbanization. The urbanization process is characterized by the expansion of the city from the core to peripheral areas which includes economic development, social, political forces and population density. Very rapid urbanization in the highly populated country like India, which changes natural land cover into urban land use, which is unavoidable. However, the study region Tiruppur is known as the knitwear capital of India that induces urban development in the region which results in the modification of the natural land cover. For understating the interaction between the natural landscape and human activities, land use and land cover (LULC) is considered as the important indicator. Research on land-use and land cover changes using remote sensing technology has a long history to evident. The advancement in the Remote Sensing and GIS techniques provide the fine resolution of data sets to proceed. Sentinel-2B imagery was chosen for this study for two main reasons one is that compare to Landsat imagery it has a high spatial resolution of 10 m and its radiometry includes three vegetation red edge bands. These two characteristics make the Sentinel-2B data appealing for LULC mapping. Different types of classification algorithms have been used to perform land use and land cover mapping. The study aims to create land use and land cover classification by making a comparison between different algorithms in Tiruppur by using Sentinel-2B satellite imagery. The commonly known classification algorithms, K-means, IsoData, support vector machines (SVMs), and maximum likelihood (ML) classification are adopted for investigation. This is followed by the selection of training pixels from the remaining classes to perform and compare different supervised learning algorithms for the first- and second-level classification in terms of accuracy rates. Accuracy was assessed through metrics derived from an error matrix, but primarily overall accuracy and kappa coefficient was used in allocating algorithm hierarchy. Finally, after the comparison, the highly accurate algorithm was suggested for the mapping of urban areas. The highest overall accuracy and kappa coefficient was produced by support vector machine (SVM) is due to the algorithm's relatively small number of complex decision boundaries. The results are helpful to understand the performance of the classification algorithm for the future studies.

1. INTRODUCTION

In 21st century the earth has been significantly transformed due to both natural and anthropogenic factors to the contentment of the human need (Foley et al., 2005). Ecosystem function and structure have been compromised by human activities, resulting in greater vulnerability of places, people, economic dynamics, and the climatic system (Kasperson and Kasperson, 2001; Ogle et al, 2017; Tyson et al, 2001). The magnitude and extent of land use/land cover (LULC) changes underway in many parts of the world (Feranec et al, 2017; Fuchs et al, 2015) are influenced by socio-economic and biophysical factors. These determinants are directly related to the functioning of local and national markets, international land external policies, as well as demographic and environmental conditions (Turner et al, 1993). However, the land use/land cover changes play a significant role in the natural processes as these dynamic changes directly disturb the equilibrium state of ecosystem i.e., impacting environment, agriculture and forests (Rahaman et al, 2017). In the early stages land use classifications were manually interpreted with extended time and difficulty in temporal analysis (Nemani and Running, 1997). In recent years, the advancement in remote sensing data products have made possible to evaluate the complex natural process (Roodposhti et al, 2020; Tripathy and Kumar, 2019). To understand the dynamics of landscape, GIS and Remote sensing are the cost effective and accurate alternative technique (Raziq et al, 2016)

it is very useful in the formulation, implementation and monitoring LULC change (Masilamani, 2012). In this context, the study region was adopted based on the diversified land use/land cover features and the various image classifying algorithm were used for classification (i.e., IsoData, K-Means, Maximum likelihood, Support Vector Machine). Here each algorithm was performed by employing high resolution satellite data of sentinel 2B. The aim of the present study is to assess the performance of classification algorithms for land use / land cover change in urban and its environs and identifying the suitable algorithm with the accuracy assessment for the study area.

2. STUDY AREA

The study area Tiruppur is known as the knitwear capital of India that induces urban development in the region which results in the modification of the natural land cover. Tiruppur Corporation is located at 10°14' N to 77°27' E and 11°20' N and 77°56' E on the banks of the Noyyal River (Fig. 1). It covers an area of 1552.98 sq.km and situated 450 kilometers southwest of the state capital Chennai and about 50 kilometers east of Coimbatore. The climate in Tiruppur is tropical with the mean maximum and minimum temperatures varying between 35 to 22 °C (95 to 72 °F). The total population of the corporation as per the 2011 census is 8, 77,778 individuals. The study area Tiruppur is selected, because almost 80 percent of the study area is covered by urban land cover.

* Corresponding author

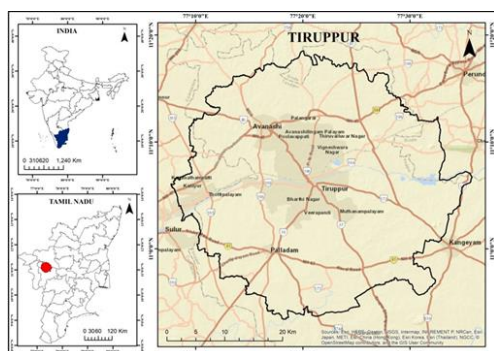


Figure 1. Study Area

However the study mainly focus on LULC classification, in specific it concentrate more on urban land use classification. It also helpful to identify the suitable classifier for urban studies

3. DATA AND METHODOLOGY

The following methodology has been applied for the sentinel 2B datasets of the study region. The sentinel 2B imagery was collected from United States Geological Survey (USGS) earth explorer on April 2020 and its spatial resolution is about 10m. The study attempts to assess the performance of the different classifiers and identification of the suitable classifier of the Land use Land cover (LULC) classification. After the image preprocessing (Atmospheric correction, Layer stacking, and Regional subset) of the sentinel 2B dataset, it further undergoes supervised and unsupervised classifications for the performance assessment and suitable classifier identification. The detailed methodological work flow of the present work is shown in the Figure 2.

3.1 Unsupervised Classification:

In unsupervised classification, the unclear image pixels are accumulated based on image clusters (Kalpana and Thanushkodi, 2010). In this Study unsupervised classification algorithms is used to detect the LULC pattern and information that was collected using specifies number of classes and iterations instead of collecting training data sets. IsoData and K-means are the unsupervised classifiers are used to perform, in IsoData evenly distributed pixels are clustered and the remaining pixels are grouped together based on defined threshold and, In K-Means the objects are developed through the gathering of similar pixels based on the center pixel of clusters (Priyadarshini et al, 2018).

3.2 Supervised classification

The training samples are collected to perform the LULC classification in supervised classifiers, each pixel in the imagery are classified into respective LULC classes with the aid of selected ROI pixels (Twisa and Buchroithner, 2019). The following three supervised algorithms are performed to undergo LULC classification in this study. The Maximum Likelihood (ML) Classifier: It assumes that each band's histogram is normally distributed and calculates the highest probability that a pixel belongs to the particular class. Support Vector Machine (SVM) is a binary classifier that performs by recognizing the paramount separating hyper plane between classes by focusing on training samples.

3.3 Post-Classification and Accuracy assessment:

The post classification process is instigated to eliminate the noises and to enhance the quality of classified output. This is

performed to confirm the accuracy of classification process. The collected ground truth ROI is used to resolve the accuracy of the classifier. The error matrix is the most pervasively used method to determine the classification accuracy (Andualem et al, 2018; Manandhar et al, 2009).

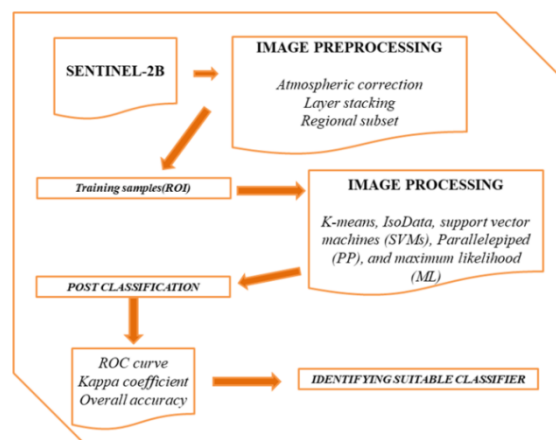


Figure 2. Methodology

4. RESULT AND DISCUSSION

The intention of the study is to analyze the performance of the classification algorithms and scrutinize the suitable classifier for the Urban Land use/ land cover (LU/LC) classification. To examine the difference in the classifiers, different types of commonly known classifiers have been used such as K-means, IsoData, support vector machines (SVM) and maximum likelihood (ML) classification. For LU/LC classification 6 classes have been categories i.e., Agricultural land (AL), Barren land (BL), Built-up (BU), Water bodies (WB), Scrub land (SL) and vegetation (VG) in each classes minimum 50 samples are extract from the source data via Region of Interest (ROI) these ROIs are taken as common input and performed on each classifiers, the results for all classifiers are shown in the figure3,4 & figure 5.

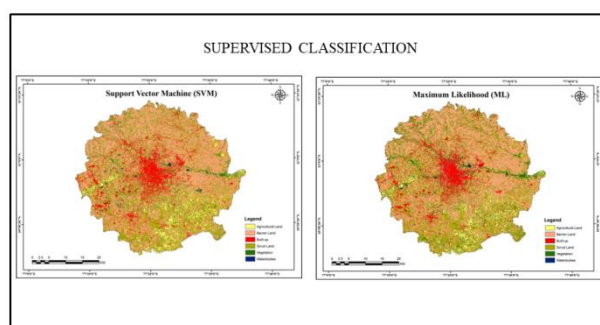


Figure 3. Supervised Classification

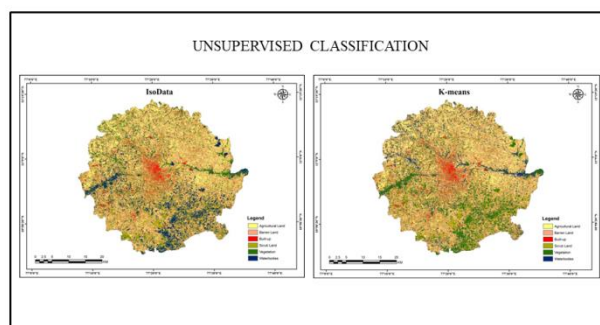


Figure 4. Unsupervised Classification

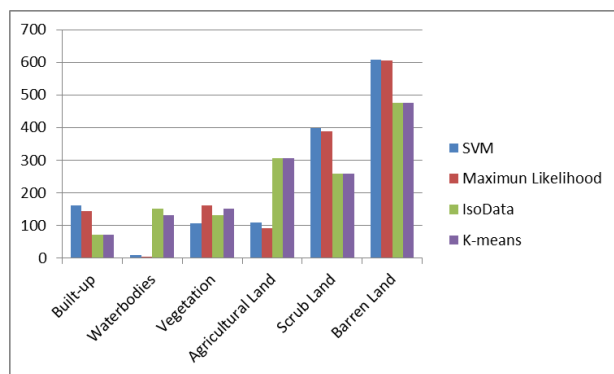


Figure 5. LULC Classes

After the image classification on each classifier, image post classifications have been done for accuracy assessment to find the suitable classifier. The various post classification parameters are Kappa coefficient, Overall accuracy and ROC curve which are shown in the Table(1)

Table 1. Kappa Coefficient & Overall Accuracy

Classification Algorithm	Kappa Coefficient	Overall Accuracy (%)
SVM	0.87	92.17
ML	0.82	88.73
IsoData	0.61	85.62
K-means	0.71	78.65

Source: Compiled by Authors

The areal extent of each class in relation with the different classifiers was computed as portrayed in the table 1. The ground truth points (GTP) have been taken out from Google earth imagery to perform the ROC curve validation part. Based on the analysis support vector machine shows the better results followed by maximum likelihood and Isodata. It is determined because of the kappa coefficient (0.878) and overall accuracy (92.17%) of SVM possess comparatively high in margin than others also SVM classified the LULC classes more precisely in relation with the satellite imagery. For overall accuracy and kappa coefficient of all the classifiers are shown in the figure 6.

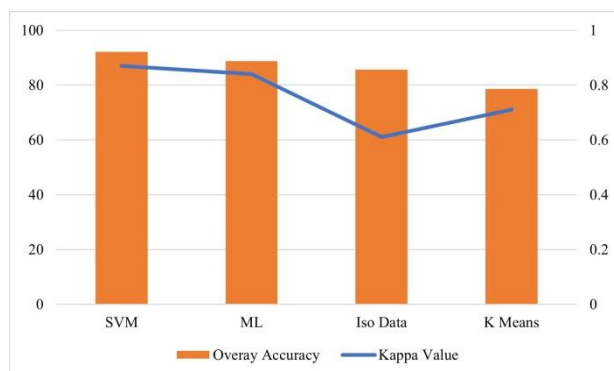


Figure 6. Kappa Coefficient & Overall Accuracy

in which SVM acquires high of kappa coefficient (0.878) and overall accuracy (92.17%) followed by ML it acquires kappa coefficient of 0.82 and over all accuracy of 88.73% then the unsupervised classifier IsoData obtains kappa coefficient of 0.78 and overall accuracy of 85.62%. The K-means is noted as the poor classifiers with 0.71 of kappa coefficient and 78.65% and of overall accuracy, respectively. The overall comprehension on the study exposed that the SVM represented

the exact classified pixels in accordance with the ground truth point, kappa coefficient and overall accuracy. To strengthen the validation of the SVM classifier ROC (Receiver Operating Characteristic) curve (Figure 7) and confusion matrix (Table 2) were generated

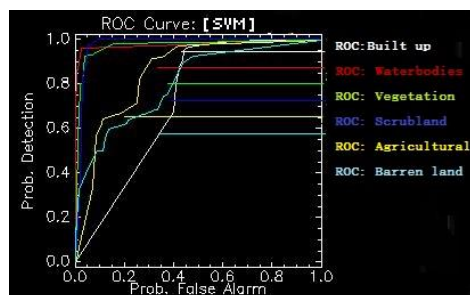


Figure 7. ROC Curve - SVM

Table 2. Confusion/Error Matrix of SVM Classifier

Class	BL	WB	VG	SL	AL
UC	0	0	0	0	0
BU	2551	3	2	0	2
WB	5	442	0	0	0
VG	5	1	199	24	7
SL	0	1	9	617	2
AL	27	2	0	8	333
BL	153	0	1	0	31
Total	2741	449	211	649	375

Source: Compiled by Authors

ROC usually denotes that if the AUC (Area under the ROC Curve) values are closer to the 0.5 the performance of the classifier is poor and if closer to 1 the classifier is excellent. Here, SVM classification showed that the AUC values of all the classes were closer to 1 and the ratio of misclassification and mixed pixels were comparatively very low which are shown in the confusion matrix. Undergoing all this analysis and validation in this study SVM classifier is recommended for LULC classification.

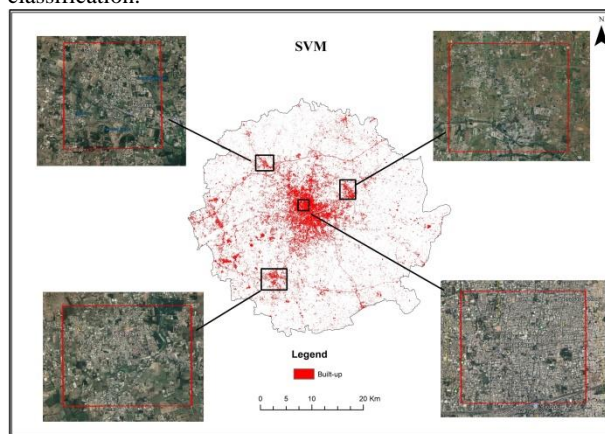


Figure 8. Google earth validation of Built up

While the study is based on the performance assessment and identification of suitable classifier for LULC, in specific the study concentrate more on urban land use classification. In this setup SVM classifier classifies overall LULC up ahead from all other classifier. In particular compare to other classifier it classify the built up more precise, it helpful to proceed the urban related models like Shannon entropy, CA markov chain model based on the SVM classification. Accordingly the present study area Tiruppur is one of the urban agglomerated cities in

Tamilnadu District. Urban sprawl and urban densification is the major problem in the past decades faced by the study area. Identification of the urban sprawl and its trends, for the sustainable urban development, LULC with precise Built up classification is inevitable. For LULC with precise built up classification SVM is the suitable classifier. The classification of the built up class by the SVM classifier is shown in the Figure 8 and the confusion matrix also helpful to validate the accuracy of the urban classification of the SVM classifier.

5. CONCLUSION

The context of the study has been assimilated in the view of identifying the better and suitable algorithm for the precise LULC classification which is the baseline and inevitable layer in every remote sensing based spatial analysis studies. The results of each LULC classes are depicting its precision in different algorithms. However, the overall accuracy and kappa coefficient values are assessed for each algorithm representing K-means (78.65%), IsoData (85.62%), support vector machines (SVMs) (92.17%) and maximum likelihood (ML) (88.73%). This illustrates that, SVM are showing better results. Especially, it is identified from the study that SVM poses the highest overall accuracy of 92.17% and specifying the significant of its efficiency in classifying the LULC features. Thus the SVM classifier is identified as significant algorithm in classifying LULC as well as the built up class in the study area, this can be applicable in assessing the urban sprawl and urban density modeling of the study area.

ACKNOWLEDGEMENT

The authors would like to thank the Department of Geography, Bharathidasan University for providing necessary support and UGC SAP DRS-II lab facilities to carry out the present work.

REFERENCES

- Rahaman, S. A., Aruchamy, S., Balasubramani, K., & Jegankumar, R. (2017). Land use/land cover changes in semi-arid mountain landscape in Southern India: A geoinformatics based Markov chain approach. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 231.
- Andualem, T. G., Belay, G., & Guadie, A. (2018). Land use change detection using remote sensing technology. *J. Earth Sci. Clim. Chang*, 9, 1-6.
- Feranec, J., Soukup, T., Taff, G., Stych, P., & Bicik, I. (1991). Overview of changes in land cover and land use in Eastern Europe. Land-cover and land-use changes in Eastern Europe after the collapse of the Soviet Union in.
- Fuchs, R., Herold, M., Verburg, P. H., Clevers, J. G., & Eberle, J. (2015). Gross changes in reconstructions of historic land cover/use for Europe between 1900 and 2010. *Global change biology*, 21(1), 299-313.
- Babykalpana, Y., & Thanushkodi, K. (2010). Classification Of Lulc Change Detection Using Remotely Sensed Data For Coimbatore City, Tamilnadu, India. *arXiv preprint arXiv:1005.4216*.
- Kasperson, J. X., & Kasperson, R. E. (2001). International workshop on vulnerability and global environmental change. A workshop summary.
- Manandhar, R., Odeh, I. O., & Ancev, T. (2009). Improving the accuracy of land use and land cover classification of Landsat data using post-classification enhancement. *Remote Sensing*, 1(3), 330-344.
- Masilamani, P. (2012). Spatio-Temporal Analysis of Urban Land Use In Ward 72 Of Madurai Corporation. In book: Orientation Programme Resource Material Publisher: Sree Venkateshwara Power Printers, Trichy.
- Nemani, R., & Running, S. (1997). Land cover characterization using multitemporal red, near-IR, and thermal-IR data from NOAA/AVHRR. *Ecological applications*, 7(1), 79-90.
- Ogle, J., Delparte, D., & Sanger, H. (2017). Quantifying the sustainability of urban growth and form through time: An algorithmic analysis of a corporation's development. *Applied Geography*, 88, 1—14.
- Priyadarshini, K. N., Kumar, M., Rahaman, S. A., & Nitheshnirmal, S. (2018). A comparative study of advanced land use/land cover classification algorithms using Sentinel-2 data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 5.
- Raziq, A., Xu, A., Li, Y., & Zhao, Q. (2016). Monitoring of land use/land cover changes and urban sprawl in Peshawar City in Khyber Pakhtunkhwa: An application of geo-information techniques using of multi-temporal satellite data. *J. Remote Sens. GIS*, 5, 174.
- Roodposhti, M. S., Hewitt, R. J., & Bryan, B. A. (2020). Towards automatic calibration of neighborhood influence in cellular automata land-use models. *Computers, Environment and Urban Systems*, 79, 101416.
- Tripathy, P., & Kumar, A. (2019). Monitoring and modelling spatio-temporal urban growth of Delhi using Cellular Automata and geoinformatics. *Cities*, 90, 52-63.
- Turner, B., Moss, R. H., & Skole, D. L. (1993). Relating land use and global land-cover change.
- Twisa, S., & Buchroithner, M. F. (2019). Land-use and land-cover (LULC) change detection in Wami River Basin, Tanzania. *Land*, 8(9), 136.
- Tyson, P., Steffen, W., Mitra, A., Fu, C., & Lebel, L. (2001). The earth system: Regional—global linkages. *Regional Environmental Change*, 2(3), 128—140.