Recent Developments in Electronics and Communication Systems
KVS Ramachandra Murthy et al. (Eds.)

doi:10.3233/ATDE221242

Image Classification of Land Use Land Cover of Bengaluru City Using Convolutional Neural Network

DR. SENTHIL KUMAR R[1] , DR.B. MEENAKSHI SUDARAM [2], Ms. RASHMI R [3]
[1] *Associate Professor, Dept. of CSE, New Horizon College of Engineering*
[2] *Professor, Dept. of CSE, NHCE,* [3] *Student, Dept. of CSE, NHCE*

Abstract. Developing countries like India is witnessing an increasing economic growth, rapid population in addition to industrialization leading to an increased rate of land use and cover. In order to better utilize the land and natural resource is essential to classify and analyse the land use and cover. Machine Learning and Deep Learning techniques are considered to be one of the effective and efficient ways for analysing and classifying the land use & cover. Here, in this paper, methodology for land use & cover classification - analysis of rural and urban regions of Bengaluru is been proposed. The proposed system's main objective is to monitor the land cover changes of Bengaluru district including its rural and urban region for classifying the land cover into its exact classes. Classification algorithms such as SVM (Support Vector Machine), RF (Random Forest), KNN (K – Nearest Neighbor) and DT (Decision Tree) are used in the preprocessing of images and model created is tested using CNN. The Landsat datasets from usgs earth explorer is used. Performance evaluation of these algorithms are done based on their accuracy rates and efficiency. The proposed system shows that CNN classifies the land cover classes efficiently because of its highest accuracy and efficiency rates when compared with other algorithms.

Keywords. CNN, Deep – Learning and Machine - Learning Techniques, Land use and cover, Landsat Dataset

1. Introduction

The forecasting of land use and cover is a crucial task due to the changing environmental conditions and alterations in the landscape. Land cover accentuates mainly on properties of nature and elements of global surface that are covered with natural bodies. Whereas, land use refers to activities of man carried out on land. Land use and cover classification and analysis has become very essential because of increase in population, economic growth, industrialization, urbanization and many more to add on. For the proper and better utilization of land and natural classification and analysis of land use and cover play a significant part. In India, cities especially like Bengaluru has been witnessing a rapid increase in growth in terms of population, development, industrialization and urbanization leading to poor or improper usage of land. This continuous development has widely affected the existence of natural and economic resources exponentially. Therefore, analyzing and observing the land cover changes and usage of land is very important for the better use of economic resources appropriately

[1] Senthil Kumar R, Dept. of CSE, NHCE, Bengaluru, India, E-mail: phdsenthilscholar@gmail.com

and also by classifying land cover into various classes like land, water, etc. helps in the proper development and utilization of land. With proper utilization of resources, land use and cover can be managed and used appropriately. Machine learning and Deep Learning techniques offers very effective and efficient means of classifying land use and cover. Machine Learning has the extent to handle wide range of giant dimension images with which it is able to classify if large amount of input sets are available. Deep Learning has strong and enhanced features in classifying the images and processing it. For the purpose of classification, ML offers techniques such as supervised, unsupervised and reinforcement. Among these, the widely used classification techniques are supervised and unsupervised. In the proposed work ML & DL algorithm uses Landsat 7 dataset of ETM i.e. Enhanced Thematic Mapper Plus which is available in USGS. These images are widely used in the change detection of land over a period and in the method proposed it is used for classification of Bengaluru region into classes like vegetation, water, barren and built-up.

The remainder of the paper is structured as follows: The literature survey is detailed in section-II; section-III discusses the preliminary design and implementation. Section-IV discusses results and in section-V of the paper conclusion is discussed.

2. Literature Survey

T Nery, et al. [1] the study's main goal was to scrutinize the performance of 6-machine learning algorithms over a time series of Landsat images for a period of 1979, 1992, 2003 and 2014 all being processed in the same manner. The algorithms used for this purpose are GBM (Stochastic Gradient Boosting), (RPART) Recursive Partitioning, Regression Trees, SVM (Support Vector Machine), RF (Random Forest), LVQ (Learning Vector Quantization) and KNN (K – Nearest Neighbor). SVM outperformed the second best classifier, RF, in terms of overall accuracy, kappa-coefficients, and marginally better fit at the individual class-level. The findings suggested that when classifying time-series imagery for land use & cover change detection, SVM or RF should be emphasized.

M. A. Z Aguilera [2] the major goal of the study was to examine different machine learning algorithms for pixel classification-of-imagery combined with Sentinel–2A & PlanetScope sensors. KNN (K – Nearest Neighbor), DT (Decision Tree), SVM (Support Vector Machine) radial base, Boosted Decision Tree, RF (Random-Forest) and ANN (Artificial Neural Network) were used for identifying four spectral indices produced from image fusion: normalized difference vegetation index, normalized difference moisture index, customized soil adjusted vegetation-index for crop health and normalized difference built-up index. The results showed that SVM radial base technique was the most accurate. A. Alem and S. Kumar [3] deep learning approaches are highlighted as a current powerful modelling methodology for extracting hidden information from large remote-sensing images for land use & cover classification in this study. Deep Learning systems capable of classifying remote sensing image data include CNN, Generative Adversarial Networks (GAN), and RNN. NWPU – Remote Sensing Image Scene Classification45, UC – Merced and Euro SAT: Sentinel – II data sets were analysed and findings were compared using various attributes and 3 distinct size remote-sensing data -sets. T. M, P. P, R. S and A. T [4-9] proposed method for the land cover categorization and detecting changes of Guntur region with high-resolution satellite images with Landsat – 7 data sets for the period 2013 and 2016. The change detection method is carried out utilizing the technique of UGG – 19 with the CNN algorithm.

Nagne Ajay D, el at. [10-14] using multispectral imagery from remotely sensed LISS – III (Linear Images Self – Scanning Sensors) , the major goal was to assess and evaluate changes in land cover and use in Aurangabad Municipal Corporation region between 2009 and 2015. The six categories of areas were classified as Water Body, Fallow, Residential Vegetation, Barren Land and Rock. For these goals, four main types of supervised algorithms are used. The Maximum Likelihood algorithm was the most accurate of them all. A. Saksena, et al. [15-20] environmental characteristics such as uneven surfaces and dirt are analysed using remote sensing imagery. The physical form of the city is linked to its functions by converting them to numerical values and classifying them. Its goal was to promote effective land usage, acquisition, and disposition in order to assure the highest & best use of land-resources. Senthil Kumar, et al. [21] proposed a deep learning methodology – for - the detection of plant diseases by performing image processing, feature extraction. Plant diseases such as early bird, black measles, leaf scorch, scab and bacterial spot was diagnosed and identified using a CNN – convolutional neural networks architecture. S Kumar, et al. [22] suggested a system to recognize birds according to their species. To network the model, an unsupervised – learning approach was utilized by utilizing deep learning, which takes the input image of bird and extracts its physical attributes such as color, wings and eyes, which are then trained as a data set and compared to the given input image.

3. Preliminary Design and Implementation

3.1 Preliminary Design

- *Architecture:*

Figure 1 depicts the proposed system architecture for the land use and cover analysis and classification where the datasets are pre – processed, features are extracted, and data sets are split into training sets and testing data. The designed model will be trained from the training data sets with selected parameters or features and then validated by using testing data sets. If the selected parameter didn't give any appropriate result, then the parameters are changed. Until the model with best accuracy is found, this process continues.

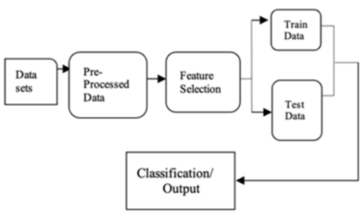


Figure 1: Architecture of Proposed System

- *Flowchart:*

Figure 2 shows a flow chart representation where the Landsat images are used as the dataset to train the classification algorithm. Before training the model, the data set is sent to pre – processing steps for the removal of unwanted and noisy values. Features needed for classification is selected. There are two types of data in the data sets: training data and testing data. Here, the model will be trained with the training data and is validated using testing data. The one giving high accuracy is selected for classification purpose.

3.2 Implementation

- *Data sets:*

The dataset used in the proposed work is Landsat 7 Enhanced Thematic Mapper Plus (ETM+) which is obtained from USGS earth explorer for the classification of land into various classes which includes built – up, water bodies, vegetation and barren using machine - learning classifiers as follows: SVM (Support Vector Machine), KNN (K – Nearest Neighbor), RF (Random Forest) and DT (Decision Tree). Here, in the proposed methodology the enormous dataset collected is divided into categories like forest, waterbodies, vegetation, industrial and residential. Figure 3 shows the one image among several images of five categories of data.

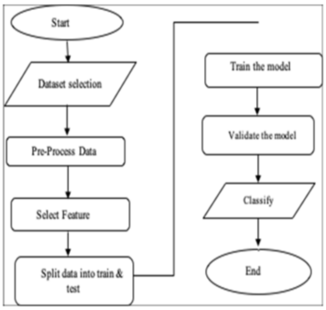


Figure 2: Flow Chart of Proposed Methodology

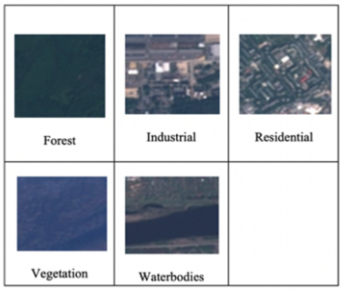


Figure 3: Five Categories of Dataset

These datasets are converted into an array format for further processing or implementation task. Figure 4 gives an array format of images.

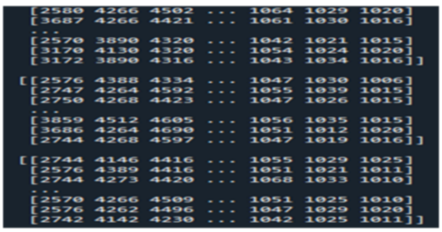


Figure 4: Dataset Converted into Array Format

- *Data Pre – Processing Method and Implementation:*

The selected dataset needs to be pre-processed in order to remove unwanted or noisy data from the dataset. The presence of unwanted or noisy data in the dataset may produce unexpected or inaccurate results when given for training the model. Interpreting from datasets becomes more difficult as the volume of data is more. In order to extract information from the dataset, several statistical procedures are required to reduce its dimensionality in a suitable manner while protecting the majority of data content.

Principal-Component-Analysis or PCA is one of the most often used approaches for this purpose with the basic goal of reducing the dimensionality of a dataset while maintaining statistical information as much as feasible. Figure 5 shows the output of the PCA technique which as taken data in the array format as input and number of PCA components is taken as 75. After PCA, normalization of dataset is done where the images whose bands are in RGB format are taken as input in order to better visualize the data. The following figure shows the outcome of each bands.

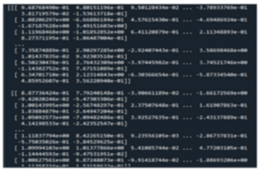


Figure 5: Outcome of PCA

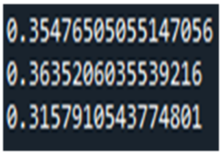


Figure 6: Normalized outcome of Each Bands

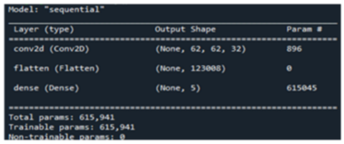


Figure 7: Model Summary of CNN

- *Implementation:*

From the references of related works, it was found that SVM, RF, KNN and DT were producing an effective outcome. Therefore, machine learning techniques, which includes algorithms such as SVM (Support Vector Machine), KNN (K – Nearest Neighbor), RF (Random Forest) and DT (Decision Tree) are used in the preprocessing of the data and building a model which is validated using CNN algorithm. Random Forest is one of supervised machine-learning -algorithm. Here, this algorithm works by building a small decision tree. The decision is based on the vote, which is been casted by building the smaller, weaker tress to form a result. K - Nearest Neighbor is a supervised machine-learning algorithm as well. K – NN has a wide range of applications, including recognition of pattern, preprocessing of data and intrusion detection. It is non – parametric, which means it does make any assumptions about the distribution of training data. When given some past data, such as training data, it divides coordinates into groups based on attribute. Another algorithm of supervised – learning called a Decision Tree is used to solve problems of classification and regression. Decision Tree can be used to train model to predict class or value of target variables by learning decision rule inferred from prior data i.e. training data. To predict a class label for a record it starts from the root of tree. Support Vector Machine is a well – known supervised -learning - technique that may be used to solve both regression and classification issues. It works by drawing a best - line or a decision - boundary via which an n – dimensional - space can be divided into various classes and new data - points may be readily - placed in the appropriate section in the following year. Hyperplane - the name given to this decision boundary. In the methodology purposed, the data sets i.e. Landsat images which is taken from usgs earth explorer is being preprocessed using pca technique and features are selected for better classification. Algorithms like Support Vector machine, Random-Forest, Decision-Tree and K – Nearest Neighbor are used here. Once this process is completed, model created is been valuated using a CNN technique. A CNN model is intended to process pixel input and is mainly used in image processing and recognition. Figure 7 shows the summary of the CNN model used during the implementation.

The model is evaluated using different measure like efficiency, accuracy, precision, recall and f-measure. Efficiency is the percentage of correct predictions multiplied by 100. According to the confusion matrix, it can be calculated as

$$\text{Efficiency} = ((TP + TN) / (TP + FN + FP + TN)) * 100$$

Where, TP = true positive, FP = false positive, FN = false negative and TN = true negative. Accuracy is the percentage of correct predictions. According to the confusion matrix, it can be calculated as

$$\text{Accuracy} = TP + TN / TP + FN + FP + TN$$

Precision is the fraction of correctly predicted positive observations among the total predicted positive observations.

$$\text{Precision} = TP / TP + FP$$

Recall is the fraction of correctly predicted positive observations among all the observations in the class.

$$\text{Recall} = TP / TP + FN$$

F-measure - the Precision and Recall criteria can be interpreted together rather than individually. To accomplish this, we consider the F-Measure values generated by the harmonic mean of the Precision and Recall columns, as the harmonic mean provides the average of two separate factors produced per unit. Therefore, F provides both the level of accuracy of the classification and how robust (less data loss) it is:

$$F - \text{Measure} = 2 * P * R / P + R$$

Where P is precision and R is Recall.

4. Results

USGS Earth Explorer was used to process the Landsat images used as a data set for land use & cover- classification. This data set is divided into two parts: training data and testing data. This training data is fed into a classification methods like SVM (Support Vector Machine), KNN (K – Nearest Neighbor), RF (Random Forest) and DT (Decision Tree) for pre-processing and the model is then validated or tested with the CNN model. Following table shows the comparison of different measures of each algorithm where the measures have been calculated using the formula mentioned above.

Table 1: Comparison of Different Measures of Each Algorithm.

Algorithms	Efficiency	Accuracy	Precision	Recall	F-measure
DT	60. 81	0.608	0.493	0.30	0.41
KNN	41. 89	0.418	0.339	0.38	0.359
RF	59.34	0.593	0.412	0.472	0.441
SVM	67. 89	0.678	0.521	0.31	0.465
CNN	78.59	0.78	0.613	0.523	0.571

From the above table it can noted that the one with highest is CNN model. Therefore, CNN is used for further classification of data sets into various classes. The confusion matrix figure 8 is provided for the same. From the above figure it can be seen that the vegetation part in the year 2010 is comparatively more than in the year 2018. From the figure it can be seen that urban area is increased drastically from 2000 to 2018. Built up has been increasing while reducing the vegetation region significantly paying way for inefficient or improper usage of economic resources and increase in land cover & use. It can be concluded that due to increase in population there is adverse effect on land cover and its use.

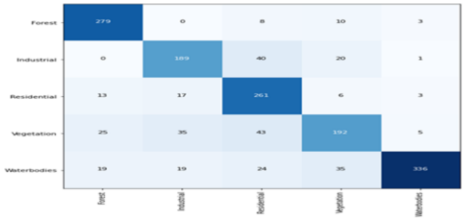
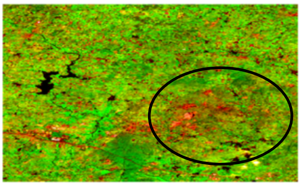


Figure 8: Confusion Matrix of CNN



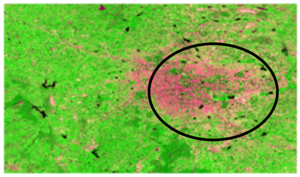


Image from 2000

Image from 2018

Figure 9: Comparison of Land in year 2000 and 2018

5. Conclusion

With the accuracy rate, the suggested method uses machine – learning and deep – learning algorithms: SVM (Support Vector Machine), KNN (K – Nearest Neighbor), RF (Random Forest) and DT (Decision Tree) and CNN. Convolutional – Neural - Network is the best algorithm. The pace of land use and cover has expanded exponentially as a result of population growth and economic growth, resulting in the exploitation of natural resources and inappropriate use of land. As a result, classification and analysis of land use and cover are critical for better natural resource management and land use.

References

[1] T Nery, et al. ‘Comparing supervised algorithms in Land Use & Land-Cover classification of a Landsat time-series’, 2016. IEEE International Geoscience & Remote Sensing Symposium pp - 5165 to 5168.

[2] M. A. Z Aguilera ‘Classification Of Land-Cover Through Machine Learning - Algorithms For Fusion Of Sentinel 2a & Planet scope Imagery’, 2020 IEEE Latin American GRSS & ISPRS Remote - Sensing Conference pp 246 to 253.

[3] Alem A, et al. ‘Deep Learning Methods for Land Cover & Land Use Classification in Remote - Sensing: A Review’, 8th International Conference on Reliability, Info com Technologies & Optimization (Trends & Future Directions) 2020, pp 903 to 908.

[4] T. M., P. P., R. S. and A. T., ‘Land Cover Change Detection Using CNN’, 3rd International conference on Electronics, Communication and Aerospace Technology, 2019, pp 791 to 794.

[5] Kussul N, et al. ‘Deep Learning Classification of Land Cover & Crop Type Using Remote Sensing Data’, in IEEE Geoscience & Remote Sensing Letters - vol. - 14, pp-778 to 782.

[6] N. Zerrouki, et al. 'A Machine Learning-Based Approach for Land Cover Change Detection Using Remote Sensing & Radiometric Measurements', in IEEE Sensors Journal, vol. 19 – pp-5843 to 5850, 15 July15, 2019.
[7] T. Vignesh, et al. 'Land Use & Land Cover -Classification Using RNNs with Shared-Layered Architecture', International Conference on Computer- Communication & Informatics, 2021, pp 1 to 6.
[8] J. Firthouse Mohammed Yashin, et al. 'Comparative Analyse of Classification Algorithm for Land use & Land cover Change - Over A Part of East Coast Region of TN & Its Environs', IEEE India Geoscience & Remote - Sensing - Symposium, 2020, pp 66 to 69.
[9] K. Kulkarni, et al. 'A comparative study of land -classification using remotely – sensed - data', 2017 International Conference on Computing - Methodologies & Communication, pp 36 to 41.
[10] Carranza-García, et al. 'A Framework for Evaluating Land Use & Land Cover Classification Using CNN'. Remote - Sens. 2019-11-274.
[11] Helber P, et al. 'Euro SAT: A Novel Dataset & Deep Learning Benchmark for Land Use & Land Cover Classification', 2017.
[12] G. J. Scott, et al. 'Training Deep NNs for Land–Cover Classification of High-Resolution Imagery', in IEEE Geoscience & Remote Sensing Letters, vol. 14, pp-549 to 553, April 2017.
[13] M. K, et. al. 'Urban land-cover classification with missing data using DCNN'. 2017 IEEE International Geoscience & Remote Sensing Symposium, pp 5161 to 5164.
[14] Pritt M, et al. 'Satellite Image Classification with Deep Learning', IEEE Applied Imagery Pattern Recognition Workshop, 2017, pp-1 to 7.
[15] Peng Dou, et al. 'Time series remote sensing images classification framework using combination of deep learning & multiple classifier system', International Journal of Applied Earth Observation & Geo information, Volume 103, 2021,102477, ISSN 0303-2434.
[16] Ienco D, et al. 'CombiningSentinel-1 & Sentinel-2 Time-Series via RNN for Object-Based Land Cover Classification', IEEE International Geoscience & Remote-Sensing -Symposium, 2019, pp 4881 to 4884.
[17] Jamali, 'Land use land cover modeling using optimized machine learning classifiers: a case study of Shiraz Iran', 2021 Model Earth Syst. Environ. 7, 1539 to 1550
[18] Rudrapal, Dhriti, and M. Subhedar. 'Land cover classification using support vector machine.' International Journal of Engineering Research & Technology (2015).
[19] Nagne, et. al. 'Land use land cover change detection by different supervised classifier on LISS - III temporal data sets', 2017 1st International Conference on Intelligent Systems& Information Management pp 68 to 71.
[20] Saksena A, et al.. 'Geo graphical Area Mapping & Classification Utilizing Multi spectral Satellite Imagery Processing Based on Machine Learning Algorithm Classifying Land based on its use for different purposes,' 2018 2nd International Conference on Electronics -Communication & Aerospace Technology, pp 1065 to 1070.
[21] Senthil, et. al. 'Plant Disease Detection &Diagnosis using Deep Learning', 2022 International Conference for Advancement in Technology, pp. 1 to 6.
[22] S. Kumar, et al. 'Analysis of Deep Learning Methods for Detection of Bird Species', 2022 Second International Conference on Artificial Intelligence & Smart Energy, pp 234 to 239.