A Review on spatial data classification on land usage and land cover using feature extraction and classification methods

[1]Gangappa Malige, [2]Dr C.Kiran Mai, [3]Dr P.Sammulal
[1] VNR VJIET ,[2] VNR VJIET, [3]JNTUH
[1] gangappa_m@vnrvjiet.in,[2] ckiranmai@vnrvjiet.in, [3]sammulalporika@gmail.com

Abstract— **Multispectral satellite images are essential source of information in the Land Usage and Land Cover (LULC) observation which is accomplished by the classification of LULC spatial data. Nevertheless, the mapping of extensive regions with spatial boundaries in spatial data is costly process .The Land Cover (LC) information describes an information source for different kinds of logical research. The classification of LC dependent on satellite information is mostly challenging assignment, and an effective classification technique is required. Most examinations have revealed that the predominant execution of image investigation on various landscape types namely urban regions, water bodies, and forest lands but, these observations have difficulties while choosing the ideal scale of segmentation, which can fall in the under or over segmentation. The different classification techniques in spatial data analysis have their own constraints, advantages, preferences and detriments, and no single strategic technique is ideal for all cases. This paper outlines and examines different soft computing methods and feature extraction strategies which are utilized for LULC spatial data classification. In view of the average error rate, exhibitions of different soft computing strategies are assessed.**

Index Terms— ***Satellite images, Soft computing, Spatial resolution, Feature extraction, Classification Methods.***

I. INTRODUCTION

The previous three decades have seen proceeding with improvements in the territory of pattern acknowledgment. Examining the algorithmic parts of pattern acknowledgment has continued by the advancement of instruments that are equipped for delivering high volumes of information, incorporating pictures with progressively better spatial and spectral resolutions [1][5]. The classification of Remote-sensing images has greatly attracted because the end results this kind of research are extensively used in many environmental, military, and financial applications. The experts in spatial and environmental research have attempted incredible endeavors in creating methodologies of classification and procedures for enhancing characterization precision [2][3]. However, it is remain a great task of characterizing spatial information into a thematic map because there exist numerous factors such as the complexity of investigation territory, remotely sensed and detected information, and image pre-processing may influence the accomplishment of classification task[5]. Though, past research is particularly worried about classifying the images [4], a audit of classification methodologies and procedures isn't accessible. In recent years, development of new characterization algorithms and systems will be exceedingly significant for controlling or choosing an appropriate technique for a particular report.

After 30 years of satellite remote sensing, the clients access complex statistical spatial data and neural calculations for both fuzzy and hard characterizations of their information.

As indicated by model of progressive structure of spatial data, the machine learning research and innovation is classified into two categories such as Surface Learning (SL) and Deep Learning (DL). The models in Surface Learning are regression methods, Support Vector Regression (SVR), neural networks. The Deep Learning (DL) is more complex and expressive strategy, and the Transform Learning (TL) is one sort of the DL. The conventional machine learning techniques need lots of training information for adjustment, which requires bunches of labor and material assets. On the other way, from the existing information, the TL can exchange information to boost the learning assignment in new condition, which does not require the indistinguishable conveyance theory as in the traditional machine learning [6][7][8][9]. Thus, by using the TL, clustering method of the Deep Convolutional Neural Network (DCNN) is transformed into task of extraction of the developed information. At first, the linear disposal model that depends on DCNN is set up to acknowledge self-learning system .This system will be trained by an example of training and learning features which are extracted from an image. In order to deal with the multispectral images, new optimal techniques are required [10]. The research and innovation in the soft computing includes accumulation of systems that abuse a resistance for vulnerability and imprecision to accomplish a minimal cost arrangement and tractability. The main aim paper is to audit and highlight the problems and challenges in the spatial data feature extraction and classification techniques utilized for LULC. This paper motivates the researchers and innovators for further research work in spatial data classification strategies. This review paper is composed as follows: Section II presents the overview of development of Landsat data. Section III gives several recent papers on LULC classification methods. The conclusion is made at the end.

II. Development of Landsat Dataset

The execution time of Landsat imagery offers analysts an opportunity to gain bits of knowledge into the patterns that are essential when checking LC changes [11]. Haack showed that Landsat pictures are utilized to take care of issues of having insufficient data on the quality of an image, particularly in growing regions. Moreover, if commercial satellite pictures are utilized, the cost of covering bigger territories can be more expensive. But, the free access of Landsat satellite images provides the opportunities to the analysts who can't manage the cost of commercial satellite images as a result of the higher costs [12][13]. Because of using these resources as a free of cost, these images solve the issue of numerous assets obliged specialists. Landsat images are enhancing continuously because the new generation satellites are being equipped with better sensors [14]. Landsat information is arranged by a system of basic frameworks situated in various regions of network of the universal co-administrators (UC) and different network stations claimed by USGS [15]. At the end of year 2016, over 57% of the spatial images captured by USGS were from this activity. But still, it is one of the best source of spatial data sets to get the spatial images. This is on-going consolidation program. However, the USGS has little information gaps because of the difficulties in changing over the information gathered from a portion of the UCs, since they are in obscure arrangements or not in good condition.

A. *Data preprocessing*

Pre-processing of spatial image includes the recognition and reconstruction of awful image data, geometric amendments, atmospheric adjustments, radiometric alignments, and topographic revisions. If the distinctive spatial subordinate data is used, information change among different sources and the quality assessment of this information are also important to be considered before that is used into machine learning classification methods. The precise geometric amendment or image enhancement of satellite data is essential for spatial information.

If a solitary-date image is used in the classification, the atmospheric corrections and radiometric rectification may not be needed. Whenever multispectral or multisensory spatial data are used, the atmospheric adjustments are required. This is particularly true whenever multisensory spatial data such as Landsat-8, radar data and SPOT sensory spatial data are coordinated for an image arrangement. An assortment of strategies, starting from basic relative alignment of image data and subtraction of dark-object in an image data depend on complex structures, have been produced for the atmospheric correction and radiometric standardization and correction.

B. *Feature Extraction in Spatial data Processing*

In any classification process, Feature Extraction (FE) is an essential step for extracting useful information. The main aim of FE in spatial image processing using machine learning methods begins with an initial set of raw spatial data and generates the optimal featured and useful data. The FE is used for the dimensionality reduction and optimal featured data for generating the spatial patterns to analyze the spatial data. Whenever the size of the input spatial data is very large, then it is transformed into a minimal set of optimal features which is represented as feature vector. The three major advantages of FE in the spatial image are as follows:

- Dimensionality reduction to get the minimal features.
- Input optimization.
- Improve accuracy with noise reduction and reduced features [16]

This feature vector will be the advanced contribution of any machine learning method [17]. In most cases, change vector investigation (CVI) is utilized for FE. The neural network system or SVM is generally used for generating the optimal features. But, in some cases principle component analysis (PCA) gives fine results with Simultaneous Self-sorting out Maps (SSOM). In many cases especially in spatial data, spectral features, and spatial features are considered from the input satellite images to find the reduced features. Local Binary Pattern (LBP) technique gives viable method for investigating surface feature strategies. It has a joined property of statistical and structural surface examination strategies [18]. The gray level co-event matrix is used to separate the formal territory boundaries between spatial structures with high spectral resolutions [19]. The Scale-invariant Feature Transform (SIFT) is one of the component oriented method for characterization of spatial data LULC. The LBP is the most broadly utilized FE technique in computer vision and pattern acknowledgment [20]. The heterogeneous features in a spatial image can be eliminated using the Partitioned feature-based classifier (PFC) and it improves the accuracy [21]. Clustering based sparse coding technique is utilized for extracting spatial features to classify multi spectral images [22].

C. *General Methods for spatial data*

The machine learning classification methods require prior knowledge for prediction and classification of image data. One can control the classification process, even though getting processed image data and training model is costly and time consuming. In the supervised machine learning model, the system of classification method is employed to train the system model with the given training data sets. In an unsupervised approach, the system model is used to group the data into different categories by using clustering algorithm. In an unsupervised classification, the human error occurs in low percentage. The spatial data such as LULC data analysis is useful for various environmental, military applications, private and public administration tasks. Remote sensing imagery is a primary source to generate land use and land cover spatial data[23][5]. The traditional classification methods may not handle uncertainty in the complex spatial data. In order to find a solution to this kind of problem, the object based classification methods are employed to classify the image data into multi objects of different sizes based on spatial-spectral features [24].

D. Land Use and Land Cover Classification data

LULC classification means segmenting multispectral satellite images into various land cover classes such as water, built-up land, farmland, forest land, bare land, etc. Different soft computing techniques used for LULC classification are SVM [25-28]. Bat algorithm [29], Fuzzy Set logic [30], Data Minimizing algorithm (EML) [31] and so on. Each technique is demonstrated based on its accuracy rate. The rate of classification accuracy depends mainly on the resolution of the image and function extraction methods. Quadratic discriminant analysis (QDA) is used in [32] for LULC classification based on spectral features.

E. Agriculture Land and Crop Identification

Soft computing techniques were used to classify various types of agricultural crops, such as coconut tree, banana tree, maize, paddy fields, wheat, etc. The classification of agricultural land and crops is more difficult than the LULC classification. Different crop species with individual color ,texture, color, etc. Based on the different texture and colors of different species, the accuracy of the soft computing techniques is used. Various soft computing techniques for the identification of agricultural land and crops are SVM[33-34], CNN, Genetic programming (GP)[35], Double crop-lands based on Continuous Wavelet Transformation (DCCWT)[36].

III. Literature Review

There are several methods suggested by researchers on the spatial images learning strategies. In this scenario, important contributions towards spatial image classification are given in the table below.

Table 1: Outline of various classification methods

Author	Methodology employed	Advantage	Performance measure
G. Zhang, et al., [25]	In this paper, a novel deep CNNs based method were implemented for very high resolution image classification[25].	For land cover types, the classification accuracy is greatly improved[25].	For land cover types, the classification accuracy is greatly improved[25].
L. Xia, et al., [26]	In this paper, an automatic classification model for remote sensing data was proposed[26].	The paper discusses automatic classification . This may reduce the labor cost for classification of land cover data	Producer's accuracy and user accuracy.
T. Wu, et al., [27]	To establish this method, an Automatic sample collection scheme for Object oriented Hierarchical Classification (AOHC) was presented[27]	The method increases the degree of automation in the land cover updating.	Producer's accuracy and user accuracy.
L. Bruzzone, and Lorenzo Carlin, [28]	This paper presented a supervised model for pixel-based system [28].	The multilevel segmentation algorithm models the neighborhood of a pixel and proper objects are generated.	Accuracy and Kappa coefficient.
H. Lu, et al., [29]	DCNN and TL (DTCLE) was proposed[29].	As the advancement in sensors, the cost of data acquisition is reduced.	Producer's accuracy, user accuracy, overall accuracy and Kappa coefficient.
J. Ma, et al., [30]	Fuzzy Local information c-means clustering algorithm (FLICM)[30] .	In FLICM algorithm[30], fuzzy factor is formulated without setting any simulated parameter[30].	Percentage Correct Classification (PCC), kappa coefficient, detection rate and false alarm rate
N. S. Mishra, et al., [31]	Fuzzy c-means (FCM) and Gustafson-Kessel clustering (GKC) are used	This is a context insensitive free model.	Missed alarms, false alarms and overall error.

Conclusion

This audit concentrated on the advancements of LULC spatial data classification strategies and deciding the most ideal methods for utilizing Landsat images in LC arrangement. The LULC spatial data characterization has kept on being a critical application, particularly with the constant presentation of advanced sensors for spatial images and the adjustment in spatial image access methods ranging from a business methods to a free data access methods. The new LULC images have enhanced characteristics namely high spatial, spectral and temporal goals. The way that LULC image can be accessed freely for any area on Earth is an additional preference. The major strength of LULC spatial data is that it describes the characterization of spatial data as the real world objects. But still, the spatial image data classification strategies have limitations in finding the suitable scale for segmentation. The challenge is to handle misclassification. This paper gives an overview of classification algorithms for LULC and also evaluates the existing methodologies by means of advantage, limitation and performance measure. Additionally, this paper evaluates the concerns faced by the existing methodologies. Still, there is much work to be done on satellite images for delivering better outcome. This paper will help the readers and researchers to understand the state-of-the-art in combination of different classifiers and motivate more meaningful works.

REFERENCES

[1] Li, Y., Xie, T., Wang, P., Wang, J., Liu, S., Zhou, X., & Zhang, X. (2018). "Joint spectral-spatial hyperspectral image classification based on hierarchical subspace switch ensemble learning algorithm". Applied Intelligence, 1-21.

[2] Li, X., Ling, F., Foody, G. M., Ge, Y., Zhang, Y., & Du, Y. (2017). "Generating a series of fine spatial and temporal resolution land cover maps by fusing coarse spatial resolution remotely sensed images and fine spatial resolution land cover maps," Remote Sensing of Environment, 196, 293-311.

[3] Knorn, J., Rabe, A., Radeloff, V. C., Kuemmerle, T., Kozak, J., & Hostert, P. (2009). "Land cover mapping of large areas using chain

classification of neighboring Landsat satellite images". Remote Sensing of Environment, 113(5), 957-964.

[4] Keshtkar, Hamidreza, Winfried Voigt, and Esmaeil Alizadeh. "Land-cover classification and analysis of change using machine-learning classifiers and multi-temporal remote sensing imagery." Arabian Journal of Geosciences 10.6 (2017), 154.

[5] M.Gangappa, Dr C.Kiran mai, Dr P.Sammulal ," ESDAS: Explorative Spatial Data Analysis Scale to predict spatial structure of landscape", International Journal of Scientific & Engineering Research 7(12), pp. 466-472,Dec 2016.

[6] M.Gangappa , C. Kiran Mai, P.Sammulal (2018) 'An NDVI based Spatial Pattern Analysis for Spatial Image Classification', International Journal of Applied Engineering Research, 13(14), pp. 11553-11558.

[7] Hengl, T., Leenaars, J. G., Shepherd, K. D., Walsh, M. G., Heuvelink, G. B., Mamo, T., ... & Wheeler, I. (2017). Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. Nutrient Cycling in Agroecosystems, 109(1), 77-102.

[8] Rajan, Suju, Joydeep Ghosh, and Melba M. Crawford. "An active learning approach to hyperspectral data classification." IEEE Transactions on Geoscience and Remote Sensing 46.4 (2008): 1231-1242.

[9] Tuia, D., Ratle, F., Pacifici, F., Kanevski, M. F., & Emery, W. J. (2009). Active learning methods for remote sensing image classification. IEEE Transactions on Geoscience and Remote Sensing, 47(7), 2218.

[10] Ienco, D., Gaetano, R., Dupaquier, C., & Maurel, P. (2017). Land cover classification via multitemporal spatial data by deep recurrent neural networks. IEEE Geoscience and Remote Sensing Letters, 14(10), 1685-1689.

[11] Turner, W., Rondinini, C., Pettorelli, N., Mora, B., Leidner, A. K., Szantoi, Z., ... & Koh, L. P. (2015). Free and open-access satellite data are key to biodiversity conservation. Biological Conservation, 182, 173-176.

[12] Mayes, Marc T., John F. Mustard, and Jerry M. Melillo. "Forest cover change in Miombo Woodlands: modeling land cover of African dry tropical forests with linear spectral mixture analysis." Remote Sensing of Environment 165 (2015): 203-215.

[13] Ernsta, C., Verhegghena, A., Bodartb, C., Mayauxb, P., de Wasseigec, C., Bararwandikad, A., ... & Kondjoh, H. K. (2010). Congo basin forest cover change estimate for 1990, 2000 and 2005 by Landsat interpretation using an automated object-based processing chain. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci, 38(6).

[14] Hansen, Matthew C., and Thomas R. Loveland. "A review of large area monitoring of land cover change using Landsat data." Remote sensing of Environment 122 (2012): 66-74.

[15] Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., ... & Roy, D. P. (2016). The global Landsat archive: Status, consolidation, and direction. Remote Sensing of Environment, 185, 271-283

[16] Li, Zechao, and Jinhui Tang. "Unsupervised feature selection via nonnegative spectral analysis and redundancy control." IEEE Transactions on Image Processing 24.12 (2015): 5343-5355.

[17] Rajadell, Olga, Pedro García-Sevilla, and Filiberto Pla. "Textural features for hyperspectral pixel classification." Iberian Conference on Pattern Recognition and Image Analysis. Springer, Berlin, Heidelberg, 2009.

[18] Eisa, M., ElGamal, A., Ghoneim, R., & Bahey, A. (2010). Local binary patterns as texture descriptors for user attitude recognition. IJCSNS International Journal of Computer Science and Network Security, 10(6), 222-229.

[19] Kuffer, M., Pfeffer, K., Sliuzas, R., & Baud, I. S. A. (2016). Extraction of slum areas from VHR imagery using GLCM variance. IEEE Journal of selected topics in applied earth observations and remote sensing, 9(5), 1830-1840.

[20] Al-Sahaf, H., Al-Sahaf, A., Xue, B., Johnston, M., & Zhang, M. (2017). Automatically evolving rotation-invariant texture image descriptors by genetic programming. IEEE Transactions on Evolutionary Computation, 21(1), 83-101.

[21] Park, Dong-Chul. "Classification of satellite images using partitioned-feature based classifier model." Information Science and Applications (ICISA), 2011 International Conference on. IEEE, 2011.

[22] Sheng, G., Yang, W., Yu, L., & Sun, H. (2012, October). Cluster structured sparse representation for high resolution satellite image classification. In Signal Processing (ICSP), 2012 IEEE 11th International Conference on (Vol. 1, pp. 693-696). IEEE.

[23] Kantakumar, L. N., & Neelamsetti, P. (2015). Multi-temporal land use classification using hybrid approach. The Egyptian Journal of Remote Sensing and Space Science, 18(2), 289-295.

[24] Bayraktar, Huseyin, and Bulent Bayram. "Fuzzy logic analysis of flood disaster monitoring and assessment of damage in se anatolia turkey." Recent Advances in Space Technologies, 2009. RAST'09. 4th International Conference on. IEEE, 2009.

[25] G. Zhang, R. Zhang, G. Zhou, and X. Jia, "Hierarchical spatial features learning with deep CNNs for very high-resolution remote sensing image classification," International Journal of Remote Sensing, pp. 1-19, 2018.

[26] L. Xia, J. Luo, W. Wang, and Z. Shen, "An Automated Approach for Land Cover Classification Based on a Fuzzy Supervised Learning Framework," Journal of the Indian Society of Remote Sensing, 42(3), 505-515, 2014.

[27] T. Wu, J. Luo, L. Xia, Z. Shen, and X. Hu, "Prior knowledge-based automatic object-oriented hierarchical classification for updating detailed land cover maps," Journal of the Indian Society of Remote Sensing, 43(4), 653-669, 2015.

[28] L. Bruzzone, and Lorenzo Carlin. "A multilevel context-based system for classification of very high spatial resolution images." IEEE transactions on Geoscience and Remote Sensing 44.9, pp. 2587-2600, 2006.

[29] H. Lu, X. Fu, C. Liu, L. G. Li, Y. X. He, and N. W. Li, "Cultivated land information extraction in UAV imagery based on deep convolutional neural network and transfer learning," Journal of Mountain Science, 14(4), 731-741, 2017.

[30] J. Ma, Maoguo Gong, and Zhiqiang Zhou. "Wavelet fusion on ratio images for change detection in SAR images." IEEE Geoscience and Remote Sensing Letters 9.6, pp. 1122-1126, 2012.

[31] N. S. Mishra, Susmita Ghosh, and Ashish Ghosh. "Fuzzy clustering algorithms incorporating local information for change detection in remotely sensed images." Applied Soft Computing 12.8, pp. 2683-2692, 2012.