2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)
University of Malaya, Kuala Lumpur, Malaysia. Sep 24-26, 2021

A Comparative Analysis of Different Land-use and Land-cover Classifiers using Hyperspectral Data

Neelam Dahiya
Chitkara University Institute of Engineering and Technology,
Chitkara University, Punjab, India.
neelam.ran@chitkara.edu.in
ORCID: 0000-0003-1839-9040

Sheifali Gupta
Chitkara University Institute of Engineering and Technology,
Chitkara University, Punjab, India.
sheifali.gupta@chitkara.edu.in
ORCID: 0000-0001-5692-418X

Sartajvir Singh
Chitkara University School of Engineering and Technology, Chitkara University,
Himachal Pradesh, India.
sartajvir.singh@chitkarauniversity.edu.in
ORCID: 0000-0002-4451-4949

2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON) | 978-1-7281-9951-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/GUCON50781.2021.9573703

***Abstract*—In the past years, various efforts have been made to extract critical information for land-use and land-cover (LULC) areas using hyperspectral imagery which was not possible with the help of multispectral imagery. Remote sensing plays an important role in providing information over large areas and in monitoring the various changes over LULC. Classification is one of the main methods which is used for the detection of changes over the earth's surface. The evaluation of different classifiers using hyperspectral imaging is essential be performed. Therefore, the main objectives of current research work are to analyze and implement the various supervised classifiers such as support vector machine (SVM), neural networks (NN), maximum likelihood classifier (MLC), and K-means as an unsupervised classifier. These classifiers have been implemented and evaluated using a hyperspectral dataset over a part of Haryana and Uttar Pradesh states, India. The results have shown that NN (89.20%) algorithm has achieved higher accuracy than other algorithms and is useful in the mapping of changes over LULC using hyperspectral imagery. This study is useful in many applications such as monitoring and mapping seasonal variations and natural resources.**

***Keywords*—Hyperspectral; support vector machine (SVM), neural networks (NN), maximum likelihood classifier (MLC); K-means.**

I. Introduction

Remote sensing plays an important role in the detection of land-use and land-cover (LULC) areas on the earth's surface which is helpful in the analysis of coverage of the particular area by forest, snow, ice, water, etc [1]–[3]. The term LULC originates from two distinct words land-use and land-cover which denotes a specific meaning. The term land-use describes how the land is used by the people for various purposes such as agriculture, transportation, and commercial use, whereas the term land-cover refers to the area which is covered by wetland, snow, water types, forest, etc [4]. On the other hand, unpredicted LULC variations due to any natural calamities lead to hazards and affect the sustainability of human activities [5], [6]. The accurate information about the LULC area helps to understand various atmospheric circulation procedures of the region which can be achieved with the help of classifiers [7].

Classification is the most common technique which is used to fetch the information from raster images and results in the formation of thematic maps [8]. Such maps are used to represent the basic features of a selected area [9]. The classification approach is further categorized into supervised and unsupervised. Supervised learning as the name indicated requires supervision for the training of the model, whereas unsupervised learning works without supervision to train the model [10]. There are various classifiers such as support vector machine (SVM) as supervised [11], maximum likelihood classifier (MLC) as supervised [12], neural network (NN) as supervised [13], and K-means as unsupervised classifier [14]. These classifiers are used to monitor the changes on the LULC with satellite imaging.

Satellite imaging technology delivers a wide range of valuable information about the earth's surface [15]. As with the constant growth of the satellite sensors, spatial resolution is also enhanced to a great extent and it may be valuable in a variety of situations [16]. Despite continuous improvements, several limitations were also reported like high-resolution satellites are required for detecting land-cover changes in the big cities, limited swath width, etc [17]. To overcome such limitations hyperspectral imaging can be significant to improve the outcomes. Hyperspectral imaging is an approach, used to collect information in numerous narrow bands from across the electromagnetic spectrum to represent a distinct variety of light energy [18]. In contrast to multispectral imaging, hyperspectral imaging provides more accurate recognition of objects [19], urban surface, plant disease, and precision agriculture [20]. Various studies reported the performance of different classifiers using a multispectral dataset [10], [21]. But there is a requirement of evaluating the well-defined and commonly used classifiers on the hyperspectral dataset.

The main focus of this work is to make a comparative analysis of various supervised i.e., SVM, MLC, NN, and unsupervised classifier i.e., K-means has been done. The accuracy assessment of each classifier has been done for the validation purpose for comparative study. This paper also portrays the use of each classifier and is organized into four main sections. The first part comprises of introduction followed by the study area and satellite dataset. The classifiers which are used for the present study are described in the third section. The fourth section includes the results and discussion of classifiers based on accuracy assessment, followed by a conclusion.

II. Study Area And Satellite Dataset

A. Study Area

The study area lies over a part of Haryana and Uttar Pradesh states, India having geographical coordinates between 30°3” N to 32°5” N in latitude and 75°3” E to 76°5” E in longitude as shown in Fig. 1. In this region, the major classes have been identified as vegetation, built-up, and water area. Vegetation is the main component of the ecosystem and covers various parts such as agriculture, forest, wetland, etc. It plays an important role in the world economy and also contributes to the cleaning of the environment [22]. So, constant monitoring and mapping are required for the management of earth resources which is possible with the help of a satellite dataset.

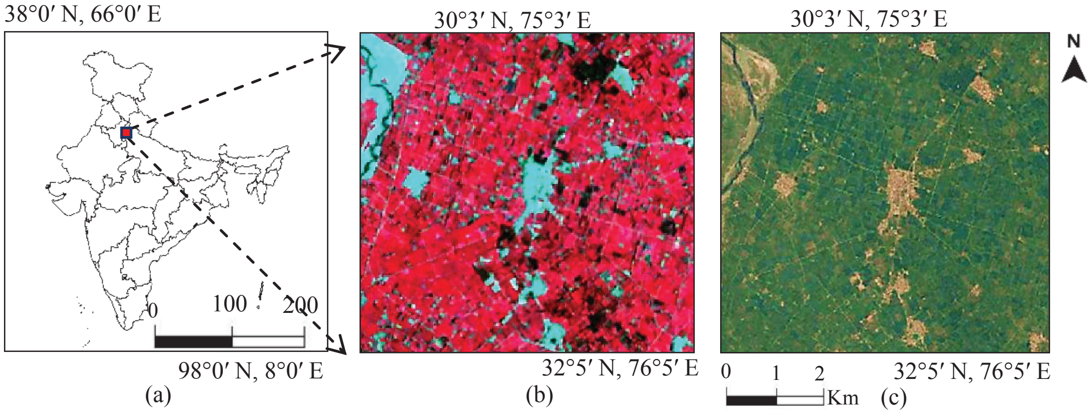

Fig. 1. Location of study site (a) Image of India (highlighted area representing selected study area) (b) False color image of the study area (07, March 2017) (c) Ground image of the study area

B. Satellite Dataset

A hyperspectral cloud-free satellite dataset was acquired on 07 March 2017. Hyperion EO-1 was used to be implemented using different classifiers. The dataset found various applications in the field of agriculture, forestry, geosciences, etc [23]. The dataset was downloaded from Unites States Geological Survey (USGS) earth explorer online web platform (https://earthexplorer.usgs.gov/). The technical specifications of the hyperspectral imaging dataset are shown in Table I. It includes the 242 spectral bands with a separation of 10 nm. It covers the wavelength range of 356-2577 nm. Due to the existence of a large number of spectral ranges, hyperspectral data offers more detailed information as compared to multispectral and helps in extracting Earth surface features.

TABLE I. TECHNICAL SPECIFICATIONS OF HYPERSPECTRAL.

S. No.	Specifications	Values/Name
1.	Dataset	Hyperion EO-1
2.	Acquisition Date	07/03/2017
2.	No. of Bands	242
3.	Wavelength	356-2577 (nm)
4.	Bandwidth	10 (nm)
5.	Spectral resolution	30 (m)
6.	Image Dimension	1001 × 3271
7.	Horizontal resolution	96 dpi
8.	Vertical resolution	96 dpi
9.	Bit depth	16
10.	File type	.TIF

III. CLASSIFICATION APPROACHES

The supervised and unsupervised classification approaches have been developed to provide an accurate assessment of classes over pixels. Supervised learning is used to predict future outcomes based on past training provided to the model on labeled data, whereas unsupervised learning does not require any training. It is a self-learning process in which the system learns the features on its own without any supervision [24]. The classification algorithm namely (a) SVM (Supervised) (b) MLC (Supervised) (c) NN (Supervised) and (d) KMC (Unsupervised) have been implemented using hyperspectral data. Each classifier is summarized below with its various advantages and disadvantages and the taxonomy of the classifier is shown as in Fig. 2.

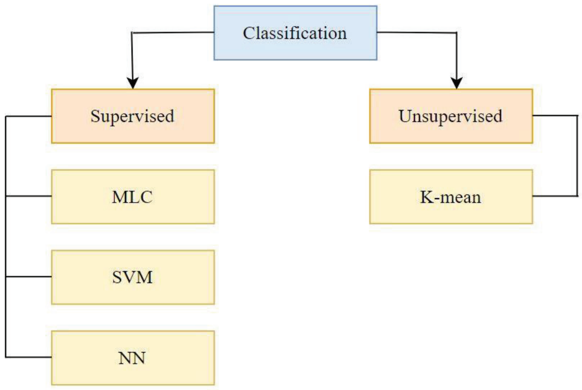

Fig. 2. Taxonomy of different classifiers considered in the present work.

A. SVM

It is a powerful technique used for classification. Instead of trying to fit the largest possible street between two classes, tries to fit as many as instances possible [25]. This algorithm is used to detect non-linear boundaries and a hyperplane is chosen to classify the objects [26]. SVM model has some excellent advantages such as (a) highly effective (b) efficient memory (c) works well with both semi-structured and unstructured data (d) effective for high dimensional data (e) less overfitting issues (f) works well for small dataset [27],[28]. However, SVM suffers from some limitations like (a) produces the crisp output in image classification i.e it usually depends upon assumptions that mentioned pixel belongs only to a single class (b) not suitable for large datasets (c) work does not well with noisy data [29].

B. MLC

It is a supervised algorithm and is one of the common methods for classification in remote sensing. In this algorithm, each pixel is allocated to that class that has the highest probability. MLC is a robust method for parameter estimation and is a widely used statistical method. It is useful for both multispectral and hyperspectral data classification. This method is also useful for urbanization mapping [21]. It also suffers from some of the drawbacks such as a time-consuming method for sample preparation. While working on multiple images, site-visit are required due to ground cover changes issues between two multitemporal datasets. Also, this method is very expensive for computation purposes [30].

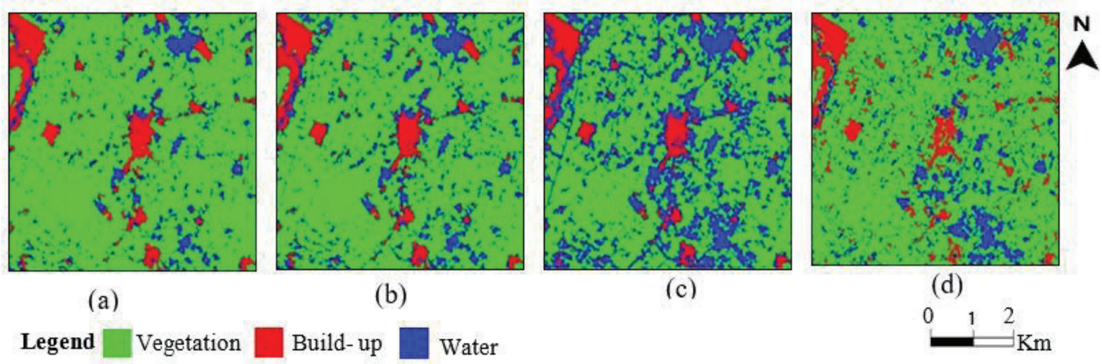

Fig. 3. Classified output using different classifiers (a) K-means (b) maximum likelihood classifier (MLC) (c) neural network (NN) (d) support vector machine (SVM).

TABLE I. ACCURACY ASSESSMENT OF DIFFERENT CLASSIFIERS.

Classifier	Classified Data	Reference Total (%)	Classified Total (%)	Correct Number (%)	Producer Accuracy (%)	User Accuracy (%)	Kappa value	Overall Accuracy (%)	Overall Kappa
K-means	Vegetation	33.6	30	81.33	72.62	81.33	0.7189		
	Built-up	38.4	40	80	83.33	80.00	0.6753	79.20	
	Water	28.0	30	76	81.43	76.00	0.6667		0.8369
MLC	Vegetation	38.4	39.2	81.63	83.33	81.63	0.7018		
	Built-up	31.2	30.0	81.33	78.21	81.33	0.7287	80.40	0.7040
	Water	30.4	30.8	77.92	78.95	77.92	0.6828		
NN	Vegetation	38	40	88	92.63	88.00	0.8065		
	Built-up	30	30	89.33	89.33	89.33	0.8476	89.20	0.6856
	Water	32	30	90.66	85.00	90.67	0.8627		
SVM	Vegetation	41.2	39.2	91.83	87.38	91.84	0.8612		
	Built-up	30.4	30.4	82.66	81.58	82.67	0.751	85.60	0.7818
	Water	28.4	28.4	80.51	87.32	80.52	0.7279		

Note MLC: maximum likelihood classifier; NN: neural network, SVM: support vector machine.

C. NN

It is a system that works like a human brain and having multiple hidden layers between input and output data. The hidden layers can be chosen according to problem complexity [31]. It stores the information from past experiences and makes it available when required. Neurons are the basic building block of artificial neural network (ANN) and are useless alone and perform magically in groups. NN requires minimal statistical training and can detect the relationship between dependent and independent variables [32]. It is also capable to work without proper knowledge. Some of the limitations of NN are (a) needs more data than other algorithms (b) hardware dependent. Instead of drawbacks, a neural network helps to solve various complex problems and in the monitoring of LULC [33].

D. K-means

It is one of the most commonly used unsupervised algorithms and is also known as an iterative algorithm. This is a clustering technique that is used to find homogeneous subgroups in the data [34]. It is an iterative algorithm and is faster than other clustering methods. Clusters help in increasing computational power and are helpful in data exploration. It is also helpful in the auto-recovery of data without any human assistance [35] [36]. This approach is beneficial when unlabeled data is required to process and produce the desired results. It also suffers from some common issues such as some clusters are difficult to predict and it is only able to handle the numerical data which directly impact the accuracy assessment[37].

IV. RESULTS AND DISCUSSION

In this paper, Hyperion EO-1 dataset is used as an input which was acquired on 07 March, 2017. The various classification algorithms such as K-means, MLC, NN, and SVM are implemented to narrate the impact on the LULC region. Hyperspectral imagery has been processed using K-means as unsupervised classification, MLC, NN, and SVM as supervised classification. During the classification various categories are explores as vegetation, built-up and water. The final classified output of different classifiers is shown in Fig. 3. Accuracy assessment is an important part of the evaluation of the performance of various classifiers. The accuracy assessment includes the following terms such as producer accuracy, user accuracy, overall accuracy, Kappa coefficient which are calculated using an error matrix [38]. The accuracy assessment is shown in Table 2. The accuracy assessment table shows that NN (89.20%) outperforms as compared to SVM (85.60%), MLC (80.40%), and K-means (79.20%) for vegetation area. Therefore, a supervised classification algorithm is preferred as compared to an unsupervised algorithm specifically for the mentioned dataset. However, the user's and producer's accuracy can be improved by making algorithms fully automatic.

V. CONCLUSION

This paper classified the data into vegetation, built-up, and water and performed the comparative analysis of various supervised and unsupervised classifiers for monitoring of LULC with the help of a hyperspectral dataset over the part of Haryana state, India. The study is capable to evaluate the Hyperion EO-1 over LULC. The results concluded that the NN achieved higher accuracy (89.20%) as compared to K-means (79.20), MLC (80.40), and SVM (85.60). The outcome of the present study shows that the NN algorithm performed well for LULC monitoring and it can be further used for another study area also. This study can be further used for forest monitoring, snow detection and crop diseases identification

ACKNOWLEDGMENT

The authors would like to thank United States Geological Survey (USGS) for providing the Hyperion EO-1 (Earth Observation) dataset for research purposes.

REFERENCES

[1] V. N. Mishra, R. Prasad, P. K. Rai, A. K. Vishwakarma, and A. Arora, "Performance evaluation of textural features in improving land use/land cover classification accuracy of heterogeneous landscape using multi-sensor remote sensing data," *Earth Sci. Informatics*, vol. 12, no. 1, pp. 71–86, 2019, doi: 10.1007/s12145-018-0369-z.

[2] A. K. Taloor *et al.*, "Land Use Land Cover Dynamics Using Remote Sensing and GIS Techniques in Western Doon Valley, Uttarakhand, India," in *Geoecology of Landscape Dynamics*, Springer, 2020, pp. 37–51.

[3] D. Kumar, A. K. Singh, A. K. Taloor, and D. Sen Singh, "Recessional pattern of Thelu and Swetvarn glaciers between 1968 and 2019, Bhagirathi basin, Garhwal Himalaya, India," *Quat. Int.*, May 2020, doi: 10.1016/j.quaint.2020.05.017.

[4] A. M. Dewan and Y. Yamaguchi, "Land use and land cover change in Greater Dhaka, Bangladesh: Using remote sensing to promote sustainable urbanization," *Appl. Geogr.*, vol. 29, no. 3, pp. 390–401, 2009, doi: 10.1016/j.apgeog.2008.12.005.

[5] D. M. McClung, "Avalanche character and fatalities in the high mountains of Asia," *Ann. Glaciol.*, vol. 57, no. 71, pp. 114–118, Jan. 2016, doi: 10.3189/2016AoG71A075.

[6] A. L. Srivastav and A. Kumar, "An endeavor to achieve sustainable development goals through floral waste management: A short review," *J. Clean. Prod.*, vol. 283, p. 124669, Feb. 2021, doi: 10.1016/j.jclepro.2020.124669.

[7] G. N. Vivekananda, R. Swathi, and A. V. L. N. Sujith, "Multi-temporal image analysis for LULC classification and change detection," *Eur. J. Remote Sens.*, vol. 00, no. 00, pp. 1–11, 2020, doi: 10.1080/22797254.2020.1771215.

[8] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 823–870, 2007, doi: 10.1080/01431160600746456.

[9] R. M. Fuller, G. M. Smith, and B. J. Devereux, "The characterisation and measurement of land cover change through remote sensing: Problems in operational applications?," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 4, no. 3, pp. 243–253, 2003, doi: 10.1016/S0303-2434(03)00004-7.

[10] V. Sood, S. Gupta, H. S. Gusain, and S. Singh, "Spatial and Quantitative Comparison of Topographically Derived Different Classification Algorithms Using AWiFS Data over Himalayas, India," *J. Indian Soc. Remote Sens.*, vol. 46, no. 12, pp. 1991–2002, 2018, doi: 10.1007/s12524-018-0861-4.

[11] M. Pal and G. M. Foody, "Evaluation of SVM, RVM and SMLR for accurate image classification with limited ground data," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 5, no. 5, pp. 1344–1355, 2012, doi: 10.1109/JSTARS.2012.2215310.

[12] S. Singh and R. Talwar, "Effects of topographic corrections on MODIS sensor satellite imagery of mountainous region," in *2013 International Conference on Signal Processing and Communication, ICSC 2013*, 2013, pp. 455–460, doi: 10.1109/ICSPCom.2013.6719833.

[13] S. Singh, R. K. Tiwari, V. Sood, and H. S. Gusain, "Detection and validation of spatiotemporal snow cover variability in the Himalayas using Ku-band (13.5 GHz) SCATSAT-1 data," *Int. J. Remote Sens.*, vol. 42, no. 3, pp. 805–815, Feb. 2021, doi: 10.1080/2150704X.2020.1825866.

[14] B. Simhachalam and G. Ganesan, "Performance comparison of fuzzy and non-fuzzy classification methods," *Egypt. Informatics J.*, vol. 17, no. 2, pp. 183–188, 2016, doi: 10.1016/j.eij.2015.10.004.

[15] S. Singh, R. K. Tiwari, H. S. Gusain, and V. Sood, "Potential Applications of SCATSAT-1 Satellite Sensor: A Systematic Review," *IEEE Sens. J.*, vol. 20, no. 21, pp. 12459–12471, Nov. 2020, doi: 10.1109/JSEN.2020.3002720.

[16] P. Verma, A. Raghubanshi, P. K. Srivastava, and A. S. Raghubanshi, "Appraisal of kappa-based metrics and disagreement indices of accuracy assessment for parametric and nonparametric techniques used in LULC classification and change detection," *Model. Earth Syst. Environ.*, vol. 6, no. 2, pp. 1045–1059, 2020, doi: 10.1007/s40808-020-00740-x.

[17] V. Sood, H. S. Gusain, S. Gupta, S. Singh, and S. Kaur, "Evaluation of SCATSAT-1 data for snow cover area mapping over a part of Western Himalayas," *Adv. Sp. Res.*, vol. 66, no. 11, pp. 2556–2567, Dec. 2020, doi: 10.1016/j.asr.2020.08.017.

[18] P. Mishra, M. S. M. Asaari, A. Herrero-Langreo, S. Lohumi, B. Diezma, and P. Scheunders, "Close range hyperspectral imaging of plants: A review," *Biosyst. Eng.*, vol. 164, pp. 49–67, 2017, doi: 10.1016/j.biosystemseng.2017.09.009.

[19] H. Okamoto and W. S. Lee, "Green citrus detection using hyperspectral imaging," *Comput. Electron. Agric.*, vol. 66, no. 2, pp. 201–208, 2009, doi: 10.1016/j.compag.2009.02.004.

[20] D. Caballero, R. Calvini, and J. M. Amigo, "Hyperspectral imaging in crop fields: precision agriculture," *Data Handl. Sci. Technol.*, vol. 32, pp. 453–473, 2020, doi: 10.1016/B978-0-444-63977-6.00018-3.

[21] B. Rimal, S. Rijal, and R. Kunwar, "Comparing Support Vector Machines and Maximum Likelihood Classifiers for Mapping of Urbanization," *J. Indian Soc. Remote Sens.*, vol. 48, no. 1, pp. 71–79, 2020, doi: 10.1007/s12524-019-01056-9.

[22] J. Im and J. R. Jensen, "Hyperspectral Remote Sensing of Vegetation," *Hyperspectral Remote Sens. Veg.*, vol. 6, pp. 1943–1961, 2016, doi: 10.1201/b11222.

[23] K. S. Khurshid *et al.*, "Preprocessing of EO-1 Hyperion data," *Can. J. Remote Sens.*, vol. 32, no. 2, pp. 84–97, 2006, doi: 10.5589/m06-014.

[24] P. C. Sen, M. Hajra, and M. Ghosh, *Supervised Classification Algorithms in Machine Learning: A Survey and Review*, vol. 937. Springer Singapore, 2020.

[25] C. Gold and P. Sollich, "Model selection for support vector machine classification," *Neurocomputing*, vol. 55, no. 1–2, pp. 221–249, 2003, doi: 10.1016/S0925-2312(03)00375-8.

[26] M. Pal and P. M. Mather, "Support vector machines for classification in remote sensing," *Int. J. Remote Sens.*, vol. 26, no. 5, pp. 1007–1011, 2005, doi: 10.1080/01431160512331314083.

[27] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, pp. 247–259, 2011, doi: 10.1016/j.isprsjprs.2010.11.001.

[28] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, 2004, doi: 10.1109/TGRS.2004.831865.

[29] S. Karamizadeh, S. M. Abdullah, M. Halimi, J. Shayan, and M. J. Rajabi, "Advantage and drawback of support vector machine functionality," *I4CT 2014 - 1st Int. Conf. Comput. Commun. Control Technol. Proc.*, no. I4ct, pp. 63–65, 2014, doi: 10.1109/I4CT.2014.6914146.

[30] J. Hogland, N. Billor, and N. Anderson, "Comparison of standard maximum likelihood classification and polytomous logistic regression used in remote sensing," *Eur. J. Remote Sens.*, vol. 46, no. 1, pp. 623–640, 2013, doi: 10.5721/EuJRS20134637.

[31] P. M. Atkinson and A. R. L. Tatnall, "Introduction neural networks in remote sensing," *Int. J. Remote Sens.*, vol. 18, no. 4, pp. 699–709, 1997, doi: 10.1080/014311697218700.

[32] A. Sharma, X. Liu, X. Yang, and D. Shi, "A patch-based convolutional neural network for remote sensing image classification," *Neural Networks*, vol. 95, pp. 19–28, 2017, doi: 10.1016/j.neunet.2017.07.017.

[33] L. Zhang, K. Wu, Y. Zhong, and P. Li, "A new sub-pixel mapping algorithm based on a BP neural network with an observation model," *Neurocomputing*, vol. 71, no. 10–12, pp. 2046–2054, 2008, doi: 10.1016/j.neucom.2007.08.033.

[34] Z. Lv, T. Liu, C. Shi, J. A. Benediktsson, and H. Du, "Novel Land Cover Change Detection Method Based on k-Means Clustering and Adaptive Majority Voting Using Bitemporal Remote Sensing Images," *IEEE Access*, vol. 7, no. c, pp. 34425–34437, 2019, doi: 10.1109/ACCESS.2019.2892648.

[35] S. J. Phillips, "Acceleration of k-means and related clustering algorithms," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2409, pp. 166–177, 2002, doi: 10.1007/3-540-45643-0_13.

[36] Bodapati S., Bandarupally H., Shaw R.N., Ghosh A. (2021) Comparison and Analysis of RNN-LSTMs and CNNs for Social Reviews Classification. In: Bansal J.C., Fung L.C.C., Simic M., Ghosh A. (eds) Advances in Applications of Data-Driven Computing. Advances in Intelligent Systems and Computing, vol 1319. Springer, Singapore. https://doi.org/10.1007/978-981-33-6919-1_4

[37] Tajammul M., Shaw R.N., Ghosh A., Parveen R. (2021) Error Detection Algorithm for Cloud Outsourced Big Data. In: Bansal J.C., Fung L.C.C., Simic M., Ghosh A. (eds) Advances in Applications of Data-Driven Computing. Advances in Intelligent Systems and Computing, vol 1319. Springer, Singapore. https://doi.org/10.1007/978-981-33-6919-1_8

[38] R. G. Congalton, “A review of assessing the accuracy of classifications of remotely sensed data,” *Remote Sens. Environ.*, vol. 37, no. 1, pp. 35–46, Jul. 1991, doi: 10.1016/0034-4257(91)90048-B.