ANALYSIS OF LAND USE LAND COVER FEATURES IN SEMI URBAN LANDSCAPE USING RANDOM FOREST

Arpitha G.A.*[1], A. L. Choodarathnakara[2], M. N. Rashmi[3], G. S. Sinchana[4] and P. K. Kumar[5]

[1, 4] Research Scholar, Dept. of E&C Engineering, Government Engineering College Kushalnagar - 571234.
EMAIL ID: arpithaga6@gmail.com, msruas.sinchanags@gmail.com
[2] Associate Professor, Dept. of E&C Engineering, Government Engineering College Karwar – 581345
EMAIL ID: choodarathnakara@gmail.com
[3] Lecturer, Dept. of Electrical and Electronics Engineering, Government Polytechnic, Mirle – 571603
EMAIL ID: rmn208@gmail.com
[5] Dept. of CSE, Visvesvaraya Technological University, PG CENTER, Mysuru – 570019
EMAIL ID: pandralli@gmail.com

Abstract: Remote sensing offers a unique environmental monitoring capability that covers extensive geographical areas, while capturing information on the Earth's atmosphere, land and oceans. The images taken are in form of pixel and process of changing it into digital images is known as Image classification. It is an effective technique to automatically discriminate classes of images with some known information. The Land Use Land Cover (LULC) describe the vegetation, water, natural surface, and cultural features on land surface. This study focuses on mapping and monitoring of LULC pattern of Virajpet Taluk, Kodagu District. The LANDSAT-8 of 2017 and 2020 data were used to extract LC map features. Random Forest is considered for classification as it is flexible for classification. From the result analysis, it is observed that OCA for the multispectral data is 51% and for fused data is 63%. Fused image gives the better classification results of LULU features than the Multispectral image. Random forest proved that is suitable for classification and produces a great accuracy most of the time.

Keywords: Remote Sensing, Land Use/Land Cover, Random Forest, Landsat-8 Satellite Imagery.

Introduction

Remote sensing is process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation from a distance. The remotely sensed image is collected, which help researchers to use in different application such as LULC. Remote sensing systems particularly those deployed on satellites provides a repetitive view of the Earth that is invaluable to monitoring the Earth system and the effect of human activities on Earth. LC refers to the surface cover on the ground such as vegetation, urban infrastructure, water, bare soil etc. and it is the physical material at the surface of the earth. Land use is the description of how people utilize the land for the socio-economic activities. LULC changes have become a vital issue that requires immediate attention due to global environmental change.

In recent years, image fusion has been explored and grabbed more attention because of the improvement in the airborne sensor which provides the both higher spatial resolution of the pan data and lower spatial resolution of multispectral data [5]. Some of researchers have used fused

ISSN: 1005-0930

image (FS) to increase the accuracy of classification. Image fusion is the technique which is used to extract the important feature information from the input image and convert it into single image which contains more information than any input image [4]. The fused image is suitable for computer processing and visual interpretation.

The author has made an attempt to merge the multispectral information from Landsat 5 images to the aerial images of the same year and month with small difference in the acquisition time of Geek cities. Unsuperviseda Pixel-based method and ISODATA technique is used for classification and experiment result showed built surfaces are about 20% underestimated, while the open surfaces are about 6% over-estimated. In case of the original MS images, the estimate of built surfaces is approximately 31% over-estimated, while the open surfaces are about 9% underestimated. Hence, FS image showed the result that is much closer to reality [2].

The researcher has demonstrated the spatial and temporal adaptive reflectance fusion model (STARFM), Bayesian maximum entropy (BME) and modified quantile–quantile adjustment (MQQA)- BME, algorithms and its performance for the image merging and fusing. The outcomes of comparison show that in the blue band, MQQA–BME algorithm have obtained higher predictionthan STARFM. MQQA-BME and BME has higher performance rate and MQQA-BME is suitable and can be implemented for application such as air quality monitoring and assessment related to epidemiology [6].

The frequency domain method named Spatial Frequency Discrete Wavelet Transform (SFDWT) was implemented by author for image fusion. This technique is applied on the Pléiades Satellite images having 1:4 resolution. The outcome of SFDWT show the better performance than original DWT [1]. The Gram–Schmidt transform and weighted median filter method was used on GF-3 and Sentinel- 2A for image fusion and random forest (RF) technique is applied for land classification. The classification result showed that the accurate map can be obtained by using WMFGS fused images [3]. Hence, RF is used widely in classification and regression problems. It can build decision trees on different samples and takes their majority vote for classification and average in cases of regression and a large number of trees can make the algorithm too slow and ineffective for real-time predictions.

This paper aims to study the LU/LC dynamics of Virajpet taluk using google earth images and LANDSAT-8 by Random Forest technique. The main objective of this paper is to examine the performance of the Random Forest classifier on LULC classification and determine the changes in land. The article is organized as follows: Section I consist of introduction to Land Use/Land cover and data fusion, Section II covers the Materials and Methodology, Result Analysis is done in Section III and Section IV concludes the final work done on the research paper.

STUDY AREA

The study area considered for in present work is Virajpet Taluk. It is situated in Kodagu district, Karnataka, India. Its geographical coordinates are 12.19' 50" North, 75.80' 40" East. It has an average elevation of 909 meters (2982 feet). The location of study area is shown in Fig 1. and Table 1 gives the specification of image data products used in this work.

Fig 1. Study area: Virajpet, Kodagu, Karnataka, India.

Table 1. Details of the data products used

Sl. No	Satellite and Data type	Date of Acquisition	Spatial Resolution
1	Landsat-8	2018/10/27	30m
2	Panchromatic data	2020/08/30	15m

Proposed Methodology

Pre-Processing

Geo-referencing:

Process of assigning the coordination points to the digital image obtained from Landsat 8 to the geographical coordination system of the earth. Fig 4. shows the Landsat 8 satellite image of Virajpete Taluk.

Sub-setting of image:

The image downloaded from the satellite will cover a larger area than the required study area. Hence the larger image has to be divided into a small area called a subset. The ERDAS software is used for experimentation.

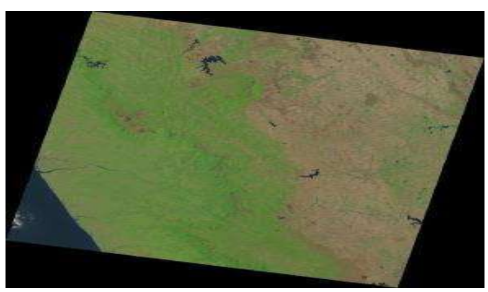

Fig 4. LANDSAT-8 Image of Virajpet

Random Forest Algorithm

The RF is basically a supervised learning algorithm. It is used for both regression and classification tasks. In this work, it is intended for classification because it's more intuitive and easier to understand. Random forest is foremost used algorithms because of its simplicity and stability. It provides higher accuracy through cross validation. While building subsets of data for trees, the word "random" comes into the picture. A subset of samples is made by randomly selecting x number of features (columns) and y number of examples (rows) from the dataset of n features and m examples. RF builds multiple decision trees and merges them together to get a more accurate and stable prediction. The RF algorithm establishes the outcome from the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of outcome.

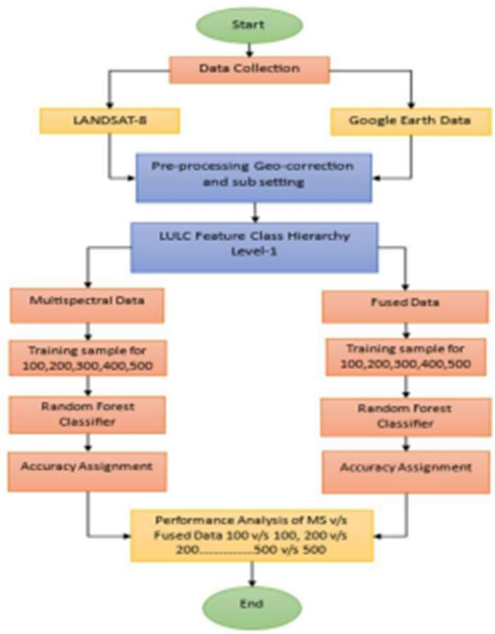

Fig 2. Methodology of Analysis LULC using Random Forest Classifier.

Accuracy assessment

RF technique is applied on fused and Multispectral data to classify LULC features such as forest, agriculture, built up, waste land, wet land, waterbodies and grassland. By increasing training samples, the overall classification accuracy (OCA) was evaluated. After conducting experiment of RF classification on Multispectral and fused data, then comparing the results of Multispectral and fused data. The confusion matrix and overall classification accuracy were used for assessment.

Result Analysis

LANDSAT-8 Multispectral Image Analysis Using Random Forest Classifier

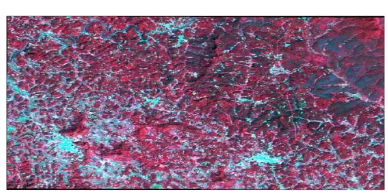

Fig 3. Original satellite image of Virajpet region

Multispectral image (MS) is a collection of a few image layers of the same scene, each of them acquired at a particular band. Multispectral imaging captures image data within specific wavelength ranges across the EM spectrum. The wavelengths may be separated by filters or detected via the use of instruments that are sensitive to particular wavelengths, including light from frequencies beyond the visible light range, i.e., infrared and ultra-violet.

Table 4. Confusion Matrix for Multispectral Image with different training sets

Confusion Matrix of Multispectral Image of TS=100

	Reference Pixels								
Classification Pixels	Classes	1	2	3	4	5	6	7	RT
	1	**0**	0	0	0	0	1	1	2
	2	0	**1**	0	1	0	0	0	2
	3	0	0	**4**	0	0	0	0	4
	4	1	0	0	**2**	0	0	1	4
	5	1	2	2	0	**0**	0	0	5
	6	0	0	0	0	0	**1**	1	2
	7	0	0	0	0	0	1	**0**	1
	CT	2	3	6	3	0	3	3	**20**

Confusion Matrix of Multispectral Image of TS=200

	Reference Pixels								
Classification Pixels	Classes	1	2	3	4	5	6	7	RT
	1	**1**	0	0	0	0	1	0	2
	2	1	**2**	1	0	0	0	0	4
	3	0	0	**6**	0	0	1	0	7
	4	4	0	2	**2**	0	2	0	10
	5	1	0	3	1	**0**	0	0	5
	6	1	0	0	0	0	**4**	0	5
	7	4	0	0	0	0	1	**1**	6
	CT	12	2	12	3	0	9	1	**39**

Confusion Matrix of Multispectral Image of TS=300

	Reference Pixels								
Classification Pixels	Classes	1	2	3	4	5	6	7	RT
	1	**3**	0	0	0	0	1	4	8
	2	0	**4**	3	0	0	1	0	8
	3	0	0	**12**	0	0	0	0	12
	4	4	0	0	**0**	0	3	1	8
	5	0	2	1	0	**0**	4	1	8
	6	0	0	0	1	0	**6**	0	7
	7	1	0	0	0	0	5	**2**	8
	CT	8	6	16	1	0	20	8	**59**

Confusion Matrix of Multispectral Image of TS=400

	Reference Pixels								
Classification Pixels	Classes	1	2	3	4	5	6	7	RT
	1	**10**	2	1	3	0	1	4	21
	2	2	**7**	5	0	1	3	1	19
	3	0	0	**15**	1	1	0	0	17
	4	5	0	2	**6**	1	2	3	19
	5	1	5	3	1	**1**	2	0	13
	6	1	0	0	3	1	**5**	5	15
	7	2	1	0	2	0	2	**7**	14
	CT	21	15	26	16	5	15	20	**118**

Confusion Matrix of Multispectral Image of TS=500

	Reference Pixels								
Classification Pixels	Classes	1	2	3	4	5	6	7	RT
	1	**2**	0	0	1	0	0	1	4
	2	0	**1**	0	1	0	0	1	3
	3	0	0	**10**	0	0	0	0	10
	4	1	0	0	**7**	0	0	1	9
	5	0	2	3	1	**0**	0	4	10
	6	2	0	0	1	0	**2**	4	9
	7	0	0	1	0	0	0	**3**	4
	CT	5	3	14	11	0	2	14	**49**

ISSN: 1005-0930

Legend: 1= Agriculture, 2= Built up, 3= Forest, 4=Grass land, 5= Waste land, 6= water Bodies, 7= Wet land, RT= Row Total, CT= Column Total

Table 4 demonstrates the confusion matrix for the RF classifier for multispectral images with 100, 200, 300, 400 and 500 training samples. The Row Total and Column total shows the Correctly Classified Pixels. Except Diagonal elements the other elements are misclassified.

4.1.1. Comparison of Overall Classification Accuracy v/s Training Samples

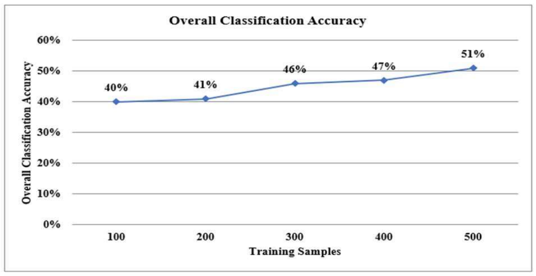

Fig 4. Comparison of overall classification accuracy v/s training samples

Fig 4. shows comparison of OCA v/s training samples OCA is the proportion of all reference pixels, which are correctly classified. For training set 100 obtained OCA is 40%. Similarly for training set 200, the obtained OCA is 41%. For the training set of 300, the obtained OCA is 46%. For the training sets of 400 and 500, the obtained OCA is 47% and 51% respectively. This comparison of overall classification accuracy v/s training samples was obtained for LANDSAT 8 Multispectral image. As the training sets were increased, the overall classification accuracy also increased.

LANDSAT-8 Fused Image Analysis Using Random Forest Classifier

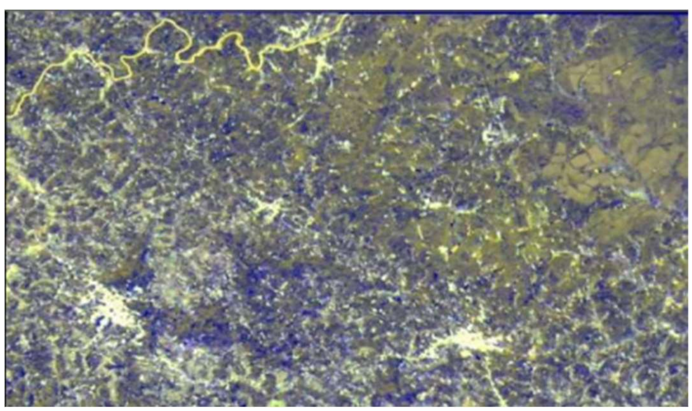

Fig 5. Original Fused Image of Virajpet

The image fusion process entails gathering crucial information from multiple images, and their inclusion into fewer images, usually a single one. This resultant image is more comprehensive and precise compared to any individual source image, containing all essential information. The goal of image fusion is to obtain a fused image that contains most significant information in all input images which were captured by different sensors from the same scene. In particular, the fusion

ISSN: 1005-0930

process should improve the contrast and keep the integrity of significant features from input images.

Table 5: Confusion Matrix for Fused Image with different training sets

Confusion Matrix of Fused Image of TS=100									
	Reference Pixels								
Classification Pixels	Classes	1	2	3	4	5	6	7	RT
	1	**0**	0	0	1	0	0	0	1
	2	0	**0**	0	0	0	0	0	0
	3	0	0	**2**	0	0	0	0	2
	4	0	0	0	**0**	0	0	0	0
	5	0	3	0	0	**1**	0	0	4
	6	0	1	0	0	0	**1**	0	2
	7	0	1	0	0	0	0	**0**	1
	CT	0	5	2	1	1	1	0	**10**

Confusion Matrix of Fused Image of TS=200									
	Reference Pixels								
Classification Pixels	Classes	1	2	3	4	5	6	7	RT
	1	**1**	0	0	0	0	2	0	3
	2	0	**0**	0	0	0	0	0	0
	3	0	0	**3**	0	0	0	0	3
	4	2	0	0	**2**	0	0	0	4
	5	0	2	0	1	**0**	0	0	3
	6	0	0	0	0	0	**3**	0	3
	7	3	0	0	1	0	0	**0**	4
	CT	6	2	3	4	0	5	0	**20**

Confusion Matrix of Fused Image of TS=300									
	Reference Pixels								
Classification Pixels	Classes	1	2	3	4	5	6	7	RT
	1	**0**	3	0	0	0	0	1	4
	2	0	**0**	0	1	0	0	0	1
	3	0	0	**7**	0	0	0	0	7
	4	0	0	1	**3**	0	0	0	4
	5	0	2	0	1	**0**	0	0	3
	6	0	1	1	0	0	**4**	1	7
	7	0	2	0	1	0	1	**0**	4
	CT	0	8	9	6	0	5	2	**30**

Confusion Matrix of Fused Image of TS=400									
	Reference Pixels								
Classification Pixels	Classes	1	2	3	4	5	6	7	RT
	1	**1**	0	0	4	0	1	1	7
	2	0	**5**	0	0	0	0	0	5
	3	0	0	**4**	0	0	0	0	4
	4	0	0	1	**6**	0	1	0	8
	5	1	2	0	1	**0**	0	0	4
	6	0	0	0	1	0	**6**	0	7
	7	3	0	1	1	0	0	**0**	5
	CT	5	7	6	13	0	8	1	**40**

Confusion Matrix of Fused Image of TS=500									
	Reference Pixels								
Classification Pixels	Classes	1	2	3	4	5	6	7	RT
	1	**2**	1	0	1	0	0	2	6
	2	0	**4**	0	0	0	0	1	5
	3	0	0	**11**	0	0	0	0	11
	4	0	0	0	**10**	0	0	0	10
	5	0	1	0	2	**0**	0	0	3
	6	0	0	0	2	0	**2**	3	7
	7	2	0	1	1	0	1	**2**	7
	CT	4	6	12	16	0	3	8	**49**

Table 5 depicts the confusion matrix for the RF technique for fused images obtained for different training samples of 100, 200, 300, 400 and 500. The table illustrates the misclassified and correctly classified points. The diagonal components show the correct classified points for different sizes of training samples.

ISSN: 1005-0930

4.2.1. Comparison of Overall Classification Accuracy v/s Training Samples

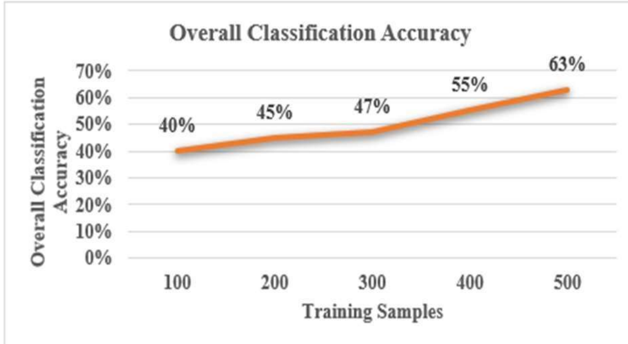

Fig 6. Comparison of overall classification accuracy v/s training samples

The comparison of Fused overall classification accuracy v/s training samples is done as shown in Fig 6. The Random Forest algorithm is applied on different datasets of LANDSAT-8. For training set 100 obtained OCA is 40%. Similarly for training set 200, the obtained OCA is 45%. For the training set of 300, the obtained OCA is 47%. For the training sets of 400 and 500, the obtained OCA is 55% and 63% respectively. This comparison of overall classification accuracy v/s training samples was obtained for LANDSAT 8 Fused image. As the training sets were increased, the overall classification accuracy also increased.

The RF algorithm is applied on different datasets of LANDSAT-8. For training set 100 obtained OCA is 40%. Similarly for training set 200, the obtained OCA is 45%. For the training set of 300, the obtained OCA is 47%. For the training sets of 400 and 500, the obtained OCA is 55% and 63% respectively. This comparison of overall classification accuracy v/s training samples was obtained for LANDSAT 8 Fused image. As the training sets were increased, the overall classification accuracy also increased.

4.2.2. Comparison of OCA v/s Training sample of Multispectral image & fused image for random forest classifier

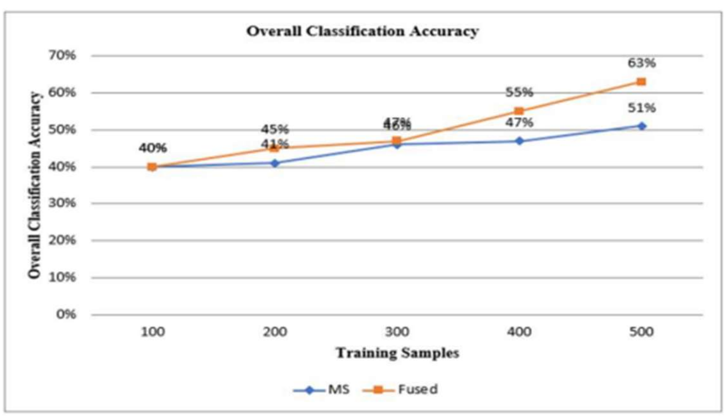

Fig 7. Comparison of OCA v/s Training Sample of Multispectral Image and Fused image for Random Forest classifier

ISSN: 1005-0930

Fig 7. illustrates the Comparison of OCA v/s Training Sample of Multispectral Image and Fused image for decision tree classifier. From the graph it is inferred that multispectral data has higher overall classification accuracy compare to Fused data. For 100-training set multispectral data has OCA of 40% while fused data have OCA of 40% and difference is 0%. For 200-training set multispectral data has OCA of 41% while fused data have OCA of 45% and difference is 4%. For 300-training set multispectral data has OCA of 46% while fused data have OCA of 47% and difference is 1%. For 400-training set multispectral data has OCA of 47% while fused data have OCA of 55% and difference is 8%. For the 500-training set multispectral data has OCA of 51% while fused data have OCA of 63% and difference is 12%.

Conclusion

The study area considered is Virajpet taluk in Kodagu district. This work was conducted to examine the accuracy of different data such as MS and fused for RF Classifier for Virajpet region. For the 100 training Set Multispectral data as OCA is 40% while for fused data has 40%. For the 200 training Set multispectral data OCA has increased by 1% while fused data is increased by 5%. For the 300 training Set multispectral data OCA has increased by 6% while fused is Increased by 7%. For Both 400 and 500 training Set multispectral data OCA has Increased by 7% and 11% while fused is Increased by 15% and 23%. It is noticed that fused image gives the better classification results of LULU features than the Multispectral image. RF is a Supervised learning algorithm; it is a flexible and easy to use ML apparoach that produces a great accuracy. It one of the predominantly used algorithms, due to of its simplicity and diversity and it also provides a better understanding of LULC features. The Thematic map of LULC features classification of Virajpet region could be useful to manage natural resources, environment and ecological regions in the Kodagu. Important issues affecting classification performance are analysed for future research work.

Acknowledgements

Acknowledgements are due towards students Chandan H. C. USN: 4GL20EC006, Banuteja V USN: 4GL17EC003, Dhanush M S USN: 4GL17EC008, Harshitha R USN: 4GL17EC014, Ramya S R USN: 4GL17EC030 of Dept of E&C Engineering, Government Engineering college, Kushalnagar, Kodagu District, Karnataka. Authors would like to graciously thank NRSC Hyderabad, KSRSAC Bangalore and Google earth for providing the data products for the study.

Author contributions

Arpitha G. A. and Choodarathnakara A. L: Conceptualization, Methodology, Data curation, Writing-Original draft preparation, Software, Validation., Sinchana G. S., Rashmi M. N., and Kumar P. K.: Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1]J. Jinju, N. Santhi, K. Ramar, & B. S. Bama, "Spatial frequency discrete wavelet transforms image fusion technique for remote sensing applications," Engineering Science and Technology, vol. 22, no. 3, pp. 715-726, 2019.

[2]D. Kaimaris, P. Patias, G. Mallinis, & C. Georgiadis, "Data Fusion of Scanned Black and White Aerial Photographs with Multispectral Satellite Images," Sci., vol. 2, no. 2, pp. 29, 2020.

[3]Y. Quan, Y. Tong, W. Feng, G. Dauphin, W. Huang, & M. Xing, "A Novel Image Fusion Method of Multi-Spectral and SAR Images for Land Cover Classification," Remote Sensing, vol. 12, no. 22, pp. 3801, 2020.

[4]D. Mishra, & B. Palkar, "Image fusion techniques: a review," International Journal of Computer Applications, vol. 130, no. 9, pp. 7-13, 2015.

[5]Y. Zhang, "Understanding image fusion. Photogramm. Eng.," Remote Sens., vol. 70, no. 6, pp. 657-661, 2004.

[6]X. Wei, N. B. Chang, & K. Bai, "A Comparative Assessment of Multisensory Data Merging and Fusion Algorithms for High-Resolution Surface Reflectance Data," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 4044-4059, 2020.

[7]J. Fortuna, H. Martens, & T. A. Johansen, "Multivariate image fusion: A pipeline for hyperspectral data enhancement," Chemometrics and Intelligent Laboratory Systems, vol. 205, pp. 104 - 097, 2020.

[8]M. Ghahremani, Y. Liu, P. Yuen, & A. Behera, "Remote sensing image fusion via compressive sensing," ISPRS journal of photogrammetry and remote sensing, vol. 152, pp. 34-48, 2019.

[9]W. Li, R. Dong, H. Fu, J. Wang, L. Yu, & P. Gong, "Integrating Google Earth imagery with Landsat data to improve 30-m resolution land cover mapping," Remote Sensing of Environment, vol. 237, pp. 111 - 563, 2020.

[10] Y. Peng, W. Li, X. Luo, J. Du, Y. Gan, & X. Gao, "Integrated fusion framework based on semicoupled sparse tensor factorization for spatio-temporal–spectral fusion of remote sensing images," Information Fusion, vol. 65, pp. 21-36.

[11] K. Rangzan, M. Kabolizadeh, D. Karimi, & S. Zareie, "Supervised cross-fusion method: a new triplet approach to fuse thermal, radar, and optical satellite data for land use classification," Environmental monitoring and assessment, vol. 191, no. 8, pp. 481, 2019.

[12] E. Shah, P. Jaya prasad, & M. E. James, "Image Fusion of SAR and Optical Images for Identifying Antarctic Ice Features," Journal of the Indian Society of Remote Sensing, vol. 47, no. 12, pp. 2113-2127, 2019.

[13] Z. Shao, J. Cai, P. Fu, L. Hu, & T. Liu, "Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product," Remote Sensing of Environment, vol. 235, pp. 111425, 2019.

[14] J. Useya, & S. Chen, "Comparative Performance Evaluation of Pixel-Level and DecisionLevel Data Fusion of Landsat 8 OLI, Landsat 7 ETM+ and Sentinel-2 MSI for Crop Ensemble Classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 11, pp. 4441-4451, 2018.

[15] H. Wang, E. Skau, H. Krim, & G. Cervone, "Fusing heterogeneous data: A case for remote sensing and social media," IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 12, pp. 6956-6968, 2018.

[16] H. Wu, S. Zhao, J. Zhang, & C. Lu, "Remote sensing image sharpening by integrating multispectral image super-resolution and convolutional sparse representation fusion," IEEE Access, vol. 7, pp. 46562-46574, 2019.

[17] S. Wu, & H. Chen, "Smart city oriented remote sensing image fusion methods based on convolution sampling and spatial transformation," Computer Communications, 2020.

[18] Y. Yang, H. Lu, S. Huang, & W. Tu, "Remote Sensing Image Fusion Based on Fuzzy Logic and Salience Measure," IEEE Geoscience and Remote Sensing Letters, 2019.

[19] F. Ye, X. Li, & X. Zhang, "FusionCNN: a remote sensing image fusion algorithm based on deep convolutional neural networks," Multimedia Tools and Applications, vol. 78, no. 11, pp. 14683-14703, 2019.

[20] N. Yokoya, P. Ghamisi, J. Xia, S. Sukhanov, R. Heremans, I. Tankoyeu, & D. Tuia, "Open data for global multimodal land use classification: Outcome of the 2017 IEEE GRSS Data Fusion Contest," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 5, pp. 1363-1377, 2018.