

CNN-based fusion and classification of SAR and Optical data

Achala Shakya , Mantosh Biswas & Mahesh Pal

To cite this article: Achala Shakya , Mantosh Biswas & Mahesh Pal (2020) CNN-based fusion and classification of SAR and Optical data, International Journal of Remote Sensing, 41:22, 8839-8861, DOI: [10.1080/01431161.2020.1783713](https://doi.org/10.1080/01431161.2020.1783713)

To link to this article: <https://doi.org/10.1080/01431161.2020.1783713>



Published online: 18 Sep 2020.



Submit your article to this journal 



View related articles 



View Crossmark data 



CNN-based fusion and classification of SAR and Optical data

Achala Shakya ^a, Mantosh Biswas  ^a and Mahesh Pal  ^b

^aComputer Engineering Department, National Institute of Technology, Kurukshetra, India; ^bCivil Engineering Department, National Institute of Technology, Kurukshetra, India

ABSTRACT

Image fusion combines the images of different spectral, spatial, multi-date, as well as radiometric data to achieve a better quality image for improved classification results. Recently, Convolution Neural Network (CNN)-based classification algorithms are extensively used for remote sensing applications. Keeping this in view, present work proposes to use CNN-based fusion and classification of Sentinel 1 (VV and VH polarization) and Sentinel 2 datasets acquired over an agricultural area near Hisar (India). For image fusion, three CNN-based approaches are used to fuse Sentinel 2 (10 m) data with VV and VH bands of Sentinel 1 data. After fusion, classification was performed using 2D-CNN classifier to judge the performance of fused images in terms of classification accuracy. Results suggest that out of the three fusion approaches, only infrared image fusion (IVF) approach performed well with the considered dataset in terms of fusion indicators and classification accuracy. Keeping in view of its better performance, this study proposes a modified IVF approach by using different image pyramid methods. Comparison of results suggests an improved performance by modified IVF approach for the fusion of Sentinel 2 and Sentinel 1 data in comparison with the original IVF approach.

ARTICLE HISTORY

Received 28 January 2020

Accepted 3 June 2020

1. Introduction

Image fusion in remote sensing is a process to merge the data acquired using different sensors/resolutions to obtain an improved image compared to those originating from individual sensor measurements (Gibril et al. 2017). Ideally, image fusion process preserves spectral characteristics of one image, whereas adding the spatial information of another image (Ghassemian et al. 2016). Image fusion is aimed at providing enhanced information and improves the interpretation of the fused output images by providing more accurate information (Rogers and Wood 1990). The main reasons behind image fusion are to sharpen the images (Meng et al. 2019), enhancing features which are not visible using a single data (Sheoran and Haack 2014), improving the missing information using different sensor images (Aschbacher and Lichtenegger 1990; Abdikan et al. 2014) and enhancing the results of geometric precision of an image (Parihar, Rathore, and Mohan 2017). Generally, image fusion utilizes three levels of fusion, namely, pixel/data level fusion, feature level fusion, and decision level fusion. All these fusion approaches

have extensively been used with remote sensing data (Pohl and Van Genderen 1998). Pixel/data level fusion operates at the image level to fuse the input images acquired from the same or different sensors and are processed together (Masi et al. 2016; Palsson, Sveinsson, and Ulfarsson 2017). Feature level fusion extracts features such as shape, neighbourhood, and the extent from the input images to achieve the desired output (Das and Ghosh 2016; Jahan and Awrangjeb 2017). In decision level fusion, the high-level information extracted from the source images is combined to form the final output (Reiche et al. 2015; Marquez et al., 2016). These image fusion approaches have been used with different types of datasets including single-sensor using temporal datasets (Karathanassi, Kolokousis, and Ioannidou 2007), multi-sensor using temporal datasets (Kulkarni and Rege 2020), single-sensor using spatial datasets (Weng, Quattrochi, and Gamba 2018), multi-sensor using spatial datasets only (Vivone 2020) and the data obtained from multiple sensors. Except for the above-mentioned fusion methods, recently other image fusion methods such as dictionary-based fusion (Lal and Anouncia 2016), Sparse Representation-based fusion (Zhu, Grohnfeldt, and Bamler 2015), Bayesian fusion (Schmitt and Zhu 2016), Component Substitution (Laben, Brower, and Co 2000; Aiazzi, Baronti, and Selva 2007), and Multi-resolution Analysis (Liu 2000; Aiazzi, Baronti, and Selva 2007) are also proposed to fuse the panchromatic/multispectral and hyperspectral data. With the developments in deep learning (DL) based algorithms in the last decades, various image fusion approaches utilizing DL algorithms for image fusion are also proposed to fuse multi-temporal, multisensor and image data acquired from different sources (Zhang, Lin, and Li 2015; Masi et al. 2016; Schmitt and Zhu 2016; Scarpa et al. 2018) which performed well with the considered dataset. The use of DL-based convolution neural networks (CNNs) such as siamese, pseudo-siamese and 2-channel siamese network have been reported for image fusion using spatial and spectral features of the remote sensing dataset (Liu et al. 2017a, 2017b, 2018; He et al. 2019; Hughes et al. 2019).

The remote sensing data such as optical and SAR (Synthetic Aperture Radar) data are also widely used for various remote sensing applications. Generally, optical data covers visible, near-infrared and short infrared wavelength region of electromagnetic radiation and their acquisition depends on the weather conditions, whereas SAR data measure physical properties of the ground objects but do not depend on the weather conditions (Tripathi, Parida, and Pandey 2019; Tripathi et al. 2020). The fusion of SAR and optical data has earlier been investigated (Schmitt, Tupin, and Zhu 2017) but the availability of free high-resolution SAR (Sentinel 1) and optical (Sentinel 2) data has allowed the research community to explore their usefulness by fusing the complimentary information of both datasets for improved image interpretation and classification results (Clerici, Valbuena Calderón, and Posada 2017; He and Yokoya et al. 2018; Benedetti et al. 2018; Hughes 2019).

The process of remote sensing image classification involves grouping the pixels having similar characteristics for different land cover types (Lillesand, Kiefer, and Chipman 2015). Over the last two decades, several state-of-the-art pixel-based classifiers are being used for remote sensing image classifications (Al-Ahmadi and Hames 2009; Mountrakis, Im, and Ogole 2011). Due to some limitations of the pixel-based classifiers, several algorithms using spectral and spatial information have also been proposed for the land cover classification and were found working well in comparison to the pixel-based classifiers (Petropoulos, Vadrevu, and Kalaitzidis 2013; Borzov et al. 2018; Ghamisi et al. 2018). This

improved performance of the spatial-spectral classifiers and the availability of the CNN-based image classification algorithms employing CNN allowed remote sensing community to use these algorithms for land cover classifications (Krizhevsky, Sutskever, and Hinton 2012; Maggiori et al. 2016; Yu, Jia, and Xu 2017; Kamilaris and Prenafeta-Boldú 2018; Zhong, Hu, and Zhou 2019; Zhang et al. 2020; Mazzia, Khaliq, and Chiaberge 2020; Li et al. 2020; Lee et al. 2020).

Recently, some studies motivated by the success of CNN in remote sensing reported the use of CNN for image fusion. Zhong et al. (2016) proposed a CNN-based image fusion algorithm for remote sensing data which was used specifically for the super-resolution of multi-spectral images and did not contribute directly to the fusion. Some other studies specifically demonstrated the use of CNN for satellite image matching (Hughes et al. 2018, 2019). An exhaustive literature review suggests that so far, few works reported the use of SAR and optical data for image fusion using CNN-based approaches.

Keeping in view of the effectiveness of CNN-based fusion and classification approaches for various remote sensing applications, this paper explores its potential for the image fusion as well as the classification of SAR (Sentinel 1) and optical (Sentinel 2) datasets over an agricultural area in India. Three deep learning-based fusion methods utilizing Laplacian pyramid were used to fuse Sentinel 1 and Sentinel 2 data. Out of three fusion approaches used in this study, the best performing approach with the used dataset was modified by implementing various pyramid methods: Gaussian Pyramid (GP), Ratio of Laplacian Pyramid (ROLP), Contrast Pyramid (CP), Gradient Pyramid (GRP), Morphological Difference Pyramid (MDP), and Filter Subtract Decimate Pyramid (FSDP) to compare the results of modified fusion approach with other approaches.

2. Background

2.1. Image fusion methods

In this study, results of the three CNN-based image fusion approaches (Liu et al. 2018, 2017a, 2017b) are used to fuse Sentinel 1 (S1) and Sentinel 2 (S2) images. A possible reason of using these three approaches for the fusion of S1 and S2 data is that they have only been used to fuse medical images, infrared-visible images, and multi-focus images, respectively.

Liu et al. (2018) proposed a four-step image fusion method (IVF) using infrared and visible images (Zhou et al. 2016), 1) generation of CNN-based weight map (W) by summarizing the image features in the input image, 2) decomposition of the image with Gaussian Laplacian Pyramid ($G\{W\}$) to represent the image features on multiple scales at different resolutions and levels, 3) calculations of fusion coefficient for image fusion. The local energy map, E_A^I and E_B^I for infrared and visible images, A and B represented as follows:

$$E_A^I = \sum_m \sum_n L\{A\}^I(i+m, j+n)^2 \quad (1)$$

$$E_B^I = \sum_m \sum_n L\{B\}^I(i+m, j+n)^2 \quad (2)$$

where $L\{A\}^l$ and $L\{B\}^l$ are pyramids of A and B images and l denotes the l^{th} decomposition level, i and j represents the pixel location in an image, m and n are the constant values (0–255). 4) Finally, reconstruction of the image by applying the Laplacian Pyramid for image fusion.

CNN-based medical image fusion method (MEDF) proposed by Liu et al. (2017a) used multi-scale processing and adaptive fusion using images acquired through Computed Tomography (CT) and Magnetic Resonance (MR) imaging process. This approach generates a weight map for integrating the pixel information from two input images so as to achieve better results with the fused image.

Liu et al. (2017b) proposed a multi-focus image fusion approach (MFF) employing deep siamese CNN (Zagoruyko and Komodakis 2015) for direct mapping between source images and focus map using dataset acquired from ImageNet dataset (<http://www.image-net.org/>). The network is trained using patches extracted from an image and blurred images for encoding to achieve better fusion performance in terms of both visual quality and the objective analysis. The initial segmentation is performed before the fusion and ‘choose-max’ strategy is applied to process the focus map, M and a threshold value is used to calculate the binary map T defined as:

$$T(i,j) = \begin{cases} 1, & M(i,j) > 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where i and j is the pixel location whose value lies between 0 and 255 and the threshold value is fixed at 0.5 to segment M to T .

The fusion is represented as follows:

$$F(i,j) = D(i,j)A(i,j) + (1 - D(i,j))B(i,j) \quad (4)$$

where D is known as the decision map and A and B are the two input images to calculate the fused image F .

The generalized architecture for pyramid-based image fusion approach can be viewed in three steps (Figure 1): the first step involves the decomposition of both SAR and optical images, after that these decomposed images are fused in the second step. Finally, a pyramidal reconstruction is used to obtain the final fused image.

All the three fusion approaches compared in this study used multi-scale fusion strategy via Laplacian pyramid only (Liu et al. 2018, 2017a, 2017b). In order to study the influence of pyramid methods on the image fusions other various state-of-the-art methods as discussed below were also implemented with the best performing image fusion approach in this study.

i. Gaussian Pyramid:

The Gaussian pyramids are also known as low-pass filtered pyramid in which a sequence of images is (Olkkinen and Pesola 1996). The main advantage of GP is that

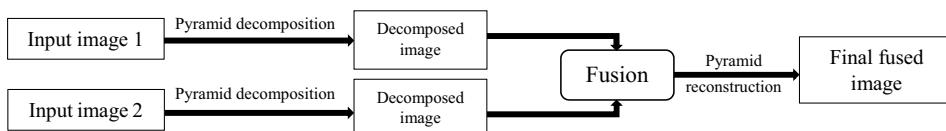


Figure 1. Generalized pyramid-based image fusion architecture.

they are useful for multi-scaled edge detection and computation of coarse-scaled images. This can be represented by the following equation:

$$x(i,j) * y(i,j) = \sum_{q=-p}^p \sum_{s=-r}^r x(q,s)f(i-q,j-s) \quad (5)$$

where $x(i,j)$ and $y(i,j)$ denotes the filter and image with the $*$ convolution operator. If $u \times v$ is the filter size, the value of p and r will be $p = (u-1)/2$ and $r = (v-1)/2$, respectively.

ii. Laplacian Pyramid:

LP is the sequence of bandpass images where each bandpass filtered image is a copy of its previous image. The copies of the bandpass images can be obtained by applying the difference operation between low-pass filtered images at successive levels of a GP (Burt and Adelson 1983; Wang and Chang 2011). LP is helpful in getting rid of blocking artefacts inside the images as well as widely useful in image restoration, composition, and enhancement. LP can be represented as follows:

$$S_i = T_i - \text{Expand}(T_{i+1}), \text{ for } 0 \leq i \leq N-1 \quad (6)$$

$$T_i = S_i + \text{Expand}(T_{i+1}), \text{ for } 0 \leq i \leq N-1 \quad (7)$$

where S_i is the sequence of the original image T for $0 \leq i \leq N-1$ levels of pyramid decomposition, T_{i+1} corresponds to low-pass filtered version, Expand represents up-sampling operation and T_i is the sequence of the images for $0 \leq i \leq N-1$ levels during pyramid decomposition.

iii. Ratio of Laplacian Pyramid:

Ratio of low-pass pyramid is the sequence of the ratio of two successive levels of the GP (Toet 1989). These pyramids are useful in encoding the absolute luminance contrasts. ROLP can be represented as follows:

$$S_i = T_i / \text{Expand}(T_{i+1}), \text{ for } 0 \leq i \leq N-1 \quad (8)$$

$$C_i = T_i / (T_{i+1}) - 1 \quad (9)$$

$$T_i = C_i \cdot \text{Expand}(T_{i+1}), \text{ for } 0 \leq i \leq N-1 \quad (10)$$

where S_i is the ratio of the original image T , T_{i+1} corresponds to low-pass filtered version, C_i represents the luminance contrast for T_i and T_{i+1} , T_i is the sequence of the images for $0 \leq i \leq N-1$ levels during pyramid decomposition.

iv. Contrast Pyramid:

CP is exactly similar to the ratio of low-pass pyramid approach. CP decomposes the input image into a series of decomposition levels having different resolution and spatial frequency (Ma, Ma, and Li 2019) which reflects the contrast information of the image. The advantage of CP is to highlight the contrast of the image and also to improve the visual interpretation of the image.

v. Gradient Pyramid:

GRP is same as the Laplacian pyramid except that GP uses a combined gradient in the GRP. At each decomposed layer of GRP, four gradient filters (horizontal, vertical, and two diagonal) are applied to obtain GRP decomposition and pyramid reconstruction is

performed in the same way as LP (Zhou, Li, and Wang 2014). GRP decomposition can be obtained as follows:

$$S_{i,n} = d_i * (T_i + w_o \cdot T_i), \text{ for } 0 \leq i \leq N-1, n = 1, 2, 3, 4 \quad (11)$$

where $S_{i,n}$ is the ratio of the original image T at successive levels for $0 \leq i \leq N-1$ levels in n direction for pyramid decomposition, d_i is the filter operator ($i = 1, 2, 3, 4$), T_i corresponds to low-pass filtered version with w_o filter.

vi. Morphological Difference Pyramid:

Morphological pyramid is obtained by applying morphological filters at each level (closing and opening transformations) to the GP which takes the difference between two neighbouring levels (Ma, Ma, and Li 2019). Morphological filters are similar to a low-pass filter but it does not alter shapes as well as locations of the objects in an image. The advantage of MDP is that it can be used to remove noise, smoothen the image, preserve geometric structure and image interpretation.

vii. Filter Subtract Decimate Pyramid:

FSDP is similar to the LP except for the use of FSD (filter, subtract, and decimate) pyramid in the place of LP. In FSD pyramid, the difference by a low-pass filter image at various image levels is directly acquired from the Gaussian image (Ma, Ma, and Li 2019). The advantage of FSD pyramid is that it is computationally more efficient than the Laplacian pyramid method as it does not perform the up-sampling step.

The main reason of using IVF, MEF and MEDF fusion approaches to fuse SAR and optical data in this study is that all the three approaches have used datasets of different modalities.

2.2. Fusion parameters

The parametric evaluation of the fused image was carried out using the objective metrics and visual assessment. Several widely used objective metrics, namely, Erreur Relative Globale Adimensionnelle de Synthese (ERGAS) (Ghassemian 2016), Spectral Angle Mapper (SAM) (Vivone et al. 2015), Relative Average Spectral Error (RASE) (Azarang and Ghassemian 2018), Universal Image Quality Index (UIQI) (Wang et al. 2004), Structural Similarity Index (SSIM) (Wang et al. 2004), Peak Signal-to-Noise Ratio (PSNR) (Wang et al. 2004), and Correlation Coefficient (CC) (Wang et al. 2004) were used to judge the quality of the fused images. The ideal value for ERGAS, SAM and RASE is zero; the reference value for UIQI and CC is 1. RASE and ERGAS parameter calculates the overall spatial and spectral distortion in the fused output, whereas SAM, RASE, and CC estimate the spectral distortion. On the other hand, SSIM reports the structural similarity whereas, UIQI estimates the spatial distortion in the fused result.

2.3. Classification method

A 2D-CNN-based classifier was used throughout in this study for all classifications using S2 data alone, layer stacked datasets consisting of, S2 with S1 (VV) and S2 with S1 (VH) polarized images as well as the fused images obtained after applying different fusion approaches. The CNN-based classifier uses ConvNet architecture (LiKamWa et al. 2016; Tran et al. 2017) which is a special type of artificial neural networks (ANNs) constructed for

feature learning and image classification tasks (LeCun et al. 1998). This architecture is highly effective for parameter sharing and learning of spatial-spectral information in the images (Yu, Jia, and Xu 2017).

CNN-based image classification consists of five different layers: input layer, convolution layer, pooling layer, fully connected layer (FC), and output layer. The full architecture of a typical CNN-based classifier used in this study comprises of an input layer, two convolution layers each combined with a pooling layer, two FC layers, followed by the output layer is shown in Figure 2.

The input layer of 2D-CNN-based classifier uses image patches of predefined sizes extracted from the satellite image using a ground reference image generated after a field visit. The work of the convolution layer is to perform the transformation of one set of feature maps into another set of feature maps using convolution with filters after passing through an activation function (Kim 2014). The input, which is fed to the network in the form of image patches of size $m \times n \times B$, where m and n are the rows and columns of the patch and B denotes the number of bands, respectively. These input patches are convolved using filters by the CNN layer to produce feature maps or activation maps. Each convolution layer is followed by a pooling layer to reduce the spatial dimensions of the activation maps leading to minimization of the computational task for the successive layers. Various pooling layers such as max-pooling and average pooling are proposed in the literature to show the importance of pooling layers (LeCun et al. 1998; Jia, Huang, and Darrell 2012; Gulcehre et al. 2014). The process of convolution and pooling is repeated based on the number of convolution layers used in the study. The activation function is also an essential parameter in designing CNN-based DL algorithms, rectified linear units (ReLU) activation function (Glorot, Bordes, and Bengio 2011) is used throughout this work. A non-linear activation function ReLU is defined by the following equation:

$$f(n) = \max(0, n) \quad (12)$$

where n is the input to the $f(n)$ (ReLU) activation function.

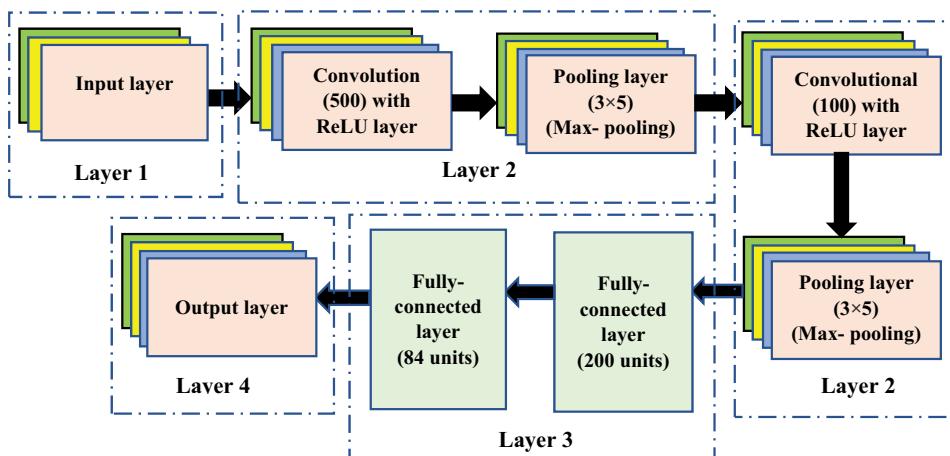


Figure 2. CNN framework used for classification.

The output of the final convolution and pooling layer acts as an input to a fully connected (FC) layer after flattening. Flattening is a process where an image is being converted to a column so that it can act as an input to a multi-layer neural network. In all these FC layers, each neuron is linked to every other present neuron. Finally, the output of the FC layer is fed as an input to the output layer, which consists of a class score using SoftMax activation function (Goodfellow, Bengio, and Courville 2016). The final result of classification is calculated using the conditional probabilities, p given below by the equation:

$$p_i = \frac{e^{z_i}}{\sum_{i=1}^F e^{z_i}} \quad (13)$$

where e is the exponential function, F is the number of classes in the output, $z = [z_1, z_2, z_3, \dots, z_F]$ is the input to the SoftMax layer. The entire network training is performed using a batch size of 200.

For the training of DL classifier, adaptive learning rate optimization algorithm (Adam; Kingma et al. 2014) with a batch size of 200 and a learning rate of 0.001 was used. The training is allowed until it reaches the pre-defined number of iterations (i.e. 100,000 in this study). To test the performance of 2D-CNN classifier, entire data were randomly divided in a way to use 75% for training and rest for the testing. Due to the availability of smaller sample sizes for few classes, oversampling was performed so as to have 1000 samples per class during training. Being an important parameter in the design of CNN, a patch size of 3×3 , found performing well with the used data, was used in this study. These patches are distributed equally into batches before feeding them to the CNN during training. For all CNN-based computations in this study, a computer with 96 GB of GPU was used with Ubuntu 14.04 operating system.

The classification results using different datasets are reported in terms of overall classification accuracy (OA) and Kappa value (K). Fused and classified images were used for comparing the results in terms of fusion indicators and classification accuracy, respectively. MATLAB and Python with Keras and Tensorflow framework were used to implement different image fusion approaches, whereas tensor flow framework was used for CNN-based image classification.

3. Dataset used and study area

The study area used in this work covers Central State farm in Hissar, Haryana (India) located at 29.2986° N and 75.7391° E. This area consists of agricultural farmland of large experimental crop fields where various seasonal crops (winter and summer) are grown mainly for seed generation to be distributed later on to the state farmers. In the study area, 12 land cover types, namely, Fallow land (1), Built-up-area (2), Dense Vegetation (3), Fenugreek (4), Fodder (5), Gram (6), Mustard (7), Oat (8), Pea (9), Sparse Vegetation (10), Spinach (11), and Wheat (12) were identified after a field visit to the study area on 6 April 2019. S1 and S2 images acquired on 23 March 2019 and 24 March 2019 were downloaded from the Copernicus open-access platform covering the study area to have no significant change in crop conditions at the time of field study. VV and VH polarized images from S1 and four bands (red (R), green (G), blue (B), and near infrared (NIR)) from

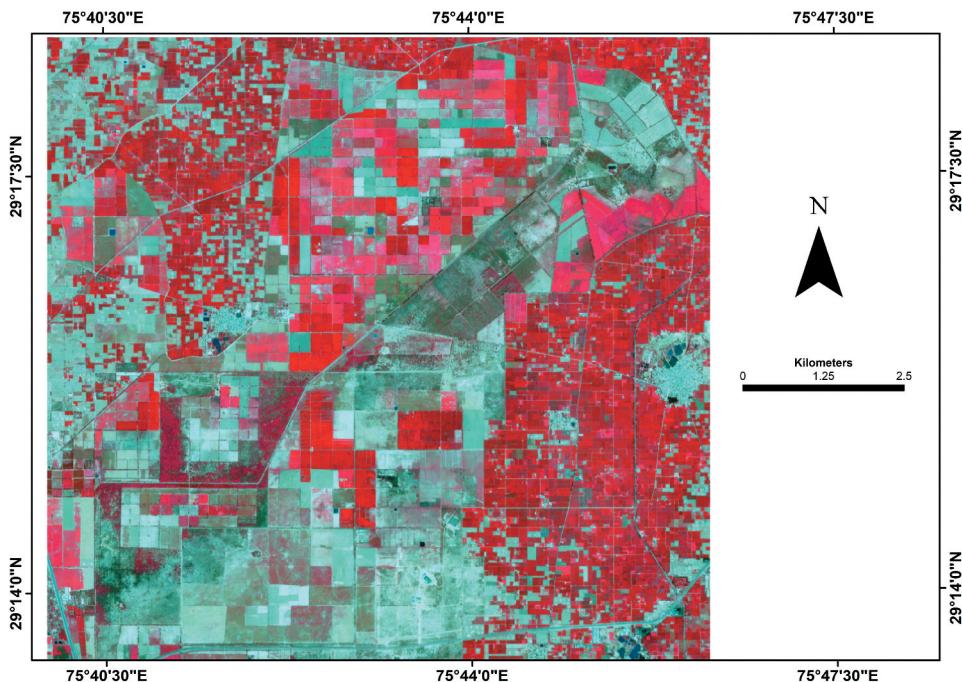


Figure 3. False colour composite image map (NIR, R and G bands) of the study area using Sentinel 2 acquired on 24 March 2019.

S2 at 10 m resolution were used for further study. False colour composite image of the study area (Sentinel 2) is provided in [Figure 3](#). Keeping in view of the large fields in the study area, the ground truth data were prepared considering the central position of the fields leading to rectangle or square shape of each field for the final classification.

4. Methodology

The workflow of the methodology used in this study is described in [Figure 4](#) and discussed below.

1.The first step involves the preprocessing of S1 and S2 data using the SNAP toolbox developed by ESA ([Figure 4](#)). After preprocessing, S1 data were resampled to 10 m resolution to perform the image fusion task with S2 data. For image fusion and classification, images of the size 964-pixel×1028-pixel were extracted from both S1 and S2 datasets covering the study area. Both VV and VH polarized bands of S1 data were used for both fusion and classification with S2 data.

2.The second step is to fuse S1 and S2 data using different CNN-based fusion approaches, namely, IVF, MEF, and MEDF approach as well as the modified IVF approach.

3.Siamese network architecture trained with ImageNet dataset (widely used to train the model for remote sensing images) was used for image fusion with all fusion approaches. This architecture was modified to suit requirements of the considered dataset in this study. For image fusion, a patch size to 32 and gamma value equal to 0.3 (factor which is multiplied to the learning rate) was used.

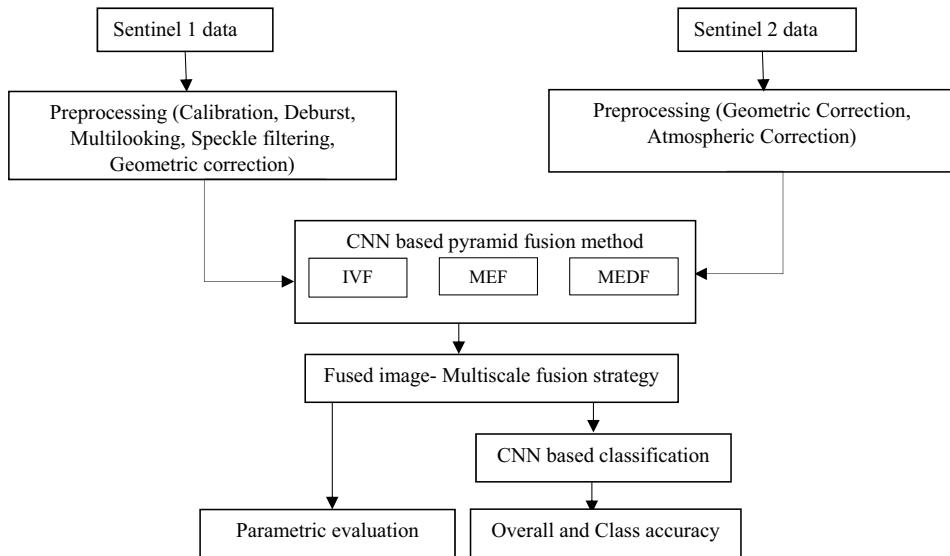


Figure 4. Methodology to fuse and classify S1 and S2 dataset.

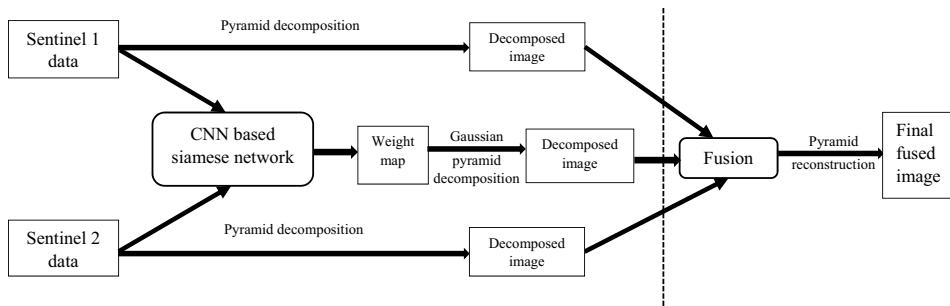


Figure 5. Detailed methodology adopted for fusion.

4. After training, the trained model was used for image fusion using different fusion approaches. To fuse both S1 and band-wise S2 data, images were passed simultaneously through channel 1 and channel 2 of the CNN-based siamese network to generate the weight map (Figure 5). Weight map thus generated was then decomposed through GP. The original input images of S1 and S2 data were also decomposed using above-mentioned pyramid approaches without passing through CNN (Figure 4). Now, these decomposed images were concatenated (to form single image) to reconstruct the pyramid images for obtaining the final fused images using 'choose max' strategy.

For the fusion of S1 and S2 data in this study, one band of S1 data (say VV) was fused with one band of S2 data using different fusion approaches. As S2 data used in this study have four different bands, thus each band is independently fused with S1 data (either VV or VH). In other words, four original bands of S2 data become four fused images as a result of the fusion of the optical bands with the SAR image (either VV or VH). These four bands are then layer stacked to form the final fused image, which is used to obtain fusion evaluation parameters as well as classification results using 2D-CNN.



5. Experimental results and discussion

5.1. Fusion results

The values of various parameters used for the evaluation of different fusion approaches after fusing S1 (VV and VH polarization) with S2 data are provided in [Table 1](#). Analysis of the results of different pyramid methods incorporated with IVF-based approach suggests better performance by ROLP-based approach (called ROLP-IVF) only, a reason of providing results result of this approach only ([Table 1](#)).

Comparison of various fusion indicators from [Table 1](#) suggests that out of IVF, MEF and MEDF fusion approach, only IVF-based fusion approach achieved better performance using VV polarized S1 data. Results in terms of fusion indicators also suggest that the proposed ROLP-IVF approach ([Table 1](#)) works well in comparison to the original IVF approach with the used dataset.

To compare the performance of different fusion approaches in terms of visual interpretation, the fused images by all fusion approaches as well as VH and VV polarized images of S1 data are provided in [Figures 6–9](#). The images obtained after fusion with both VH and VV polarization are displayed using three bands (NIR, R, and G; [Figures 6–9](#)). The rectangular box (yellow colour) in [Figures 6](#) and [8](#) is zoomed in [Figures 7](#) and [9](#), respectively, to show the effect of the image fusion on feature enhancement.

Comparison of results in terms of fusion indicators suggests that ROLP-IVF approach ([Table 1](#)) works well than the actual approach (i.e. IVF) as proposed by Liu et al. (2018). For the sake of clarity in results, the output of ROLP-IVF approach is also provided in [Figures 6–9](#).

[Figure 6](#) provides the fused images obtained by using VH polarized image of S1 data with that of S2 images using different fusion approaches. Zoomed portion (yellow rectangle in [Figure 5](#)) of a part of the fused images ([Figure 6](#)) is provided in [Figure 7](#). A comparison of zoomed portion and the fused images suggests the poor performance of MEDF fusion approach due to loss of brightness of objects, non-homogeneity (in terms of intensity) as well as the distorted texture of the features in the fused images. The results of IVF approach ([Figures 6–7\(c\)](#)) indicate rough texture of all classes in the fused image, whereas the ROLP-IVF approach ([Figures 6–7\(b\)](#)) maintains the homogeneity of the all features in the fused image. On the other hand, the MEF-based fusion approach indicates correct tone (brightness) of vegetation ([Figures 6–7\(e\)](#)) and classes appeared to be more differentiable in the fused image.

[Figures 8](#) and [9](#) provide fused images by all four fusion approaches using VV polarized image of S1 with S2 data. Visual inspection of the fused images suggests poor quality of

Table 1. Evaluation of fusion parameter.

Fusion method	Polarization	Fusion parameter						
		ERGAS	SAM	UIQI	SSIM	CC	RASE	PSNR
ROLP-IVF	VH	3.99	4.61	0.89	1.00	0.97	14.34	31.17
	VV	2.29	2.91	0.90	1.00	0.99	9.58	32.76
IVF	VH	4.14	4.89	0.86	1.00	0.96	16.99	32.27
	VV	2.57	3.17	0.90	1.00	0.98	10.35	33.31
MEDF	VH	11.39	6.36	0.63	0.10	0.81	45.49	14.25
	VV	5.75	3.92	0.85	0.73	0.95	23.02	19.84
MEF	VH	7.98	3.62	0.81	0.81	0.86	31.92	16.32
	VV	6.79	4.02	0.84	0.98	0.88	27.17	17.44

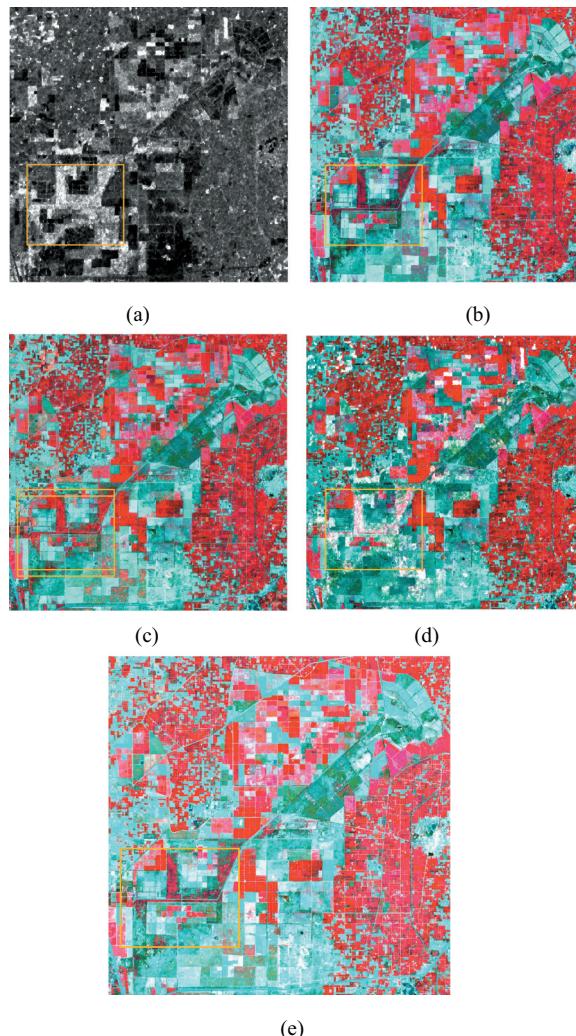


Figure 6. (a) S1 (VH), image fusion using (b) ROLP-IVF, (c) IVF, (d) MEDF and (e) MEF approaches. Yellow rectangle in figure represents the enhanced portion after fusion.

the image by MEDF approach ([Figures 8–9\(d\)](#)). Fused image obtained by IVF approach ([Figures 8–9\(c\)](#)) shows smooth texture throughout and found to preserve the shape of the sparsely vegetated areas. On the other hand, ROLP-IVF approach ([Figures 8–9\(b\)](#)) was able to maintain the homogeneity and the shape of the features throughout the region. Fused images ([Figures 8–9\(e\)](#)) using MEF approach indicates less variation in the tone of vegetations.

Based on the results in terms of various fusion indicators used to represent the quality of fused images suggest that the proposed ROLP-IVF approach performed well with the used dataset in comparison to other fusion approaches (i.e. IVF, MEF, and MEDF). Further, results also suggest better performance by VV polarized band of S1 in comparison to VH polarized band.

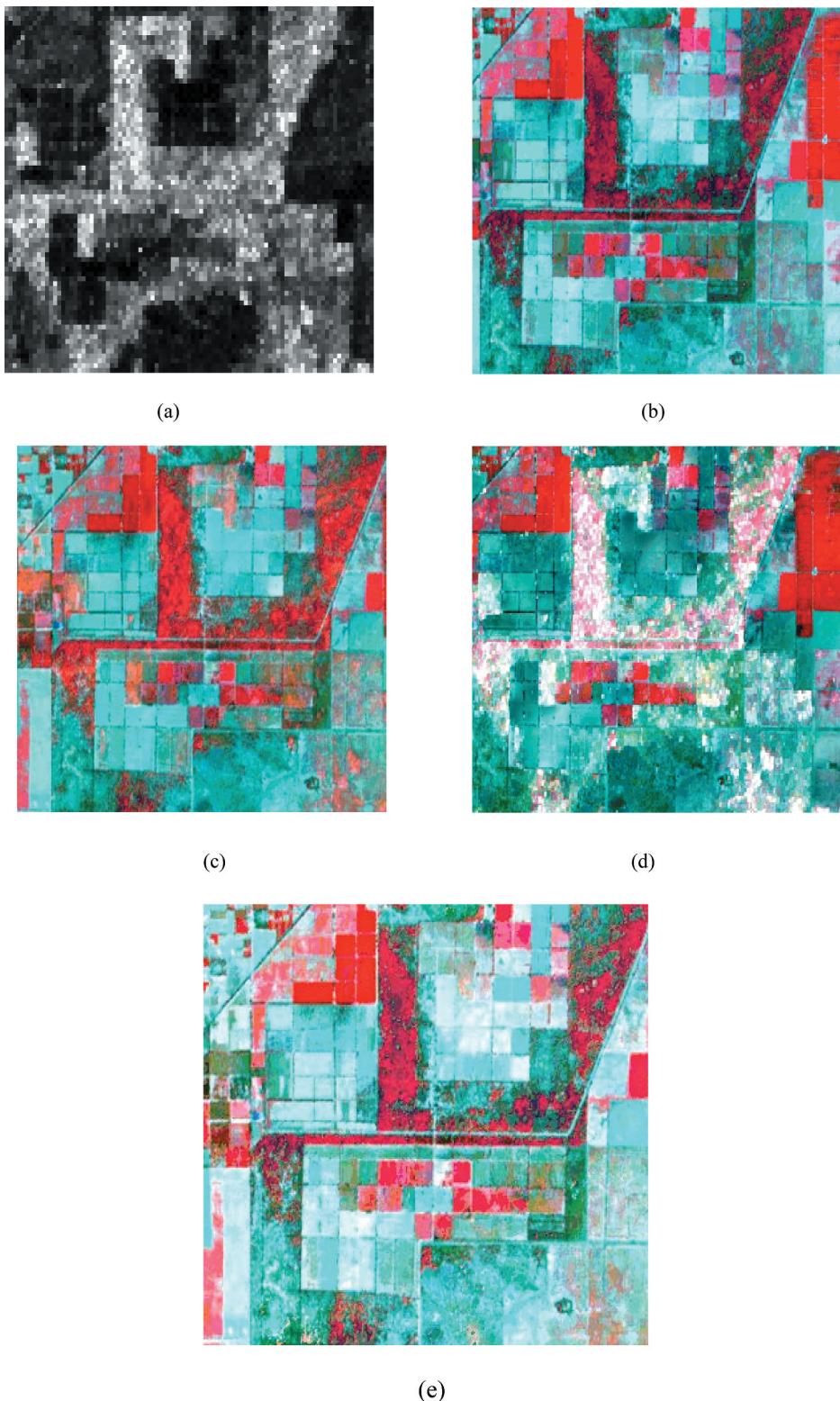


Figure 7. Zoomed Figure 5(a) S1 (VH), image fusion using (b) ROLP-IVF, (c) IVF, (d) MEDF and (e) MEF approaches.

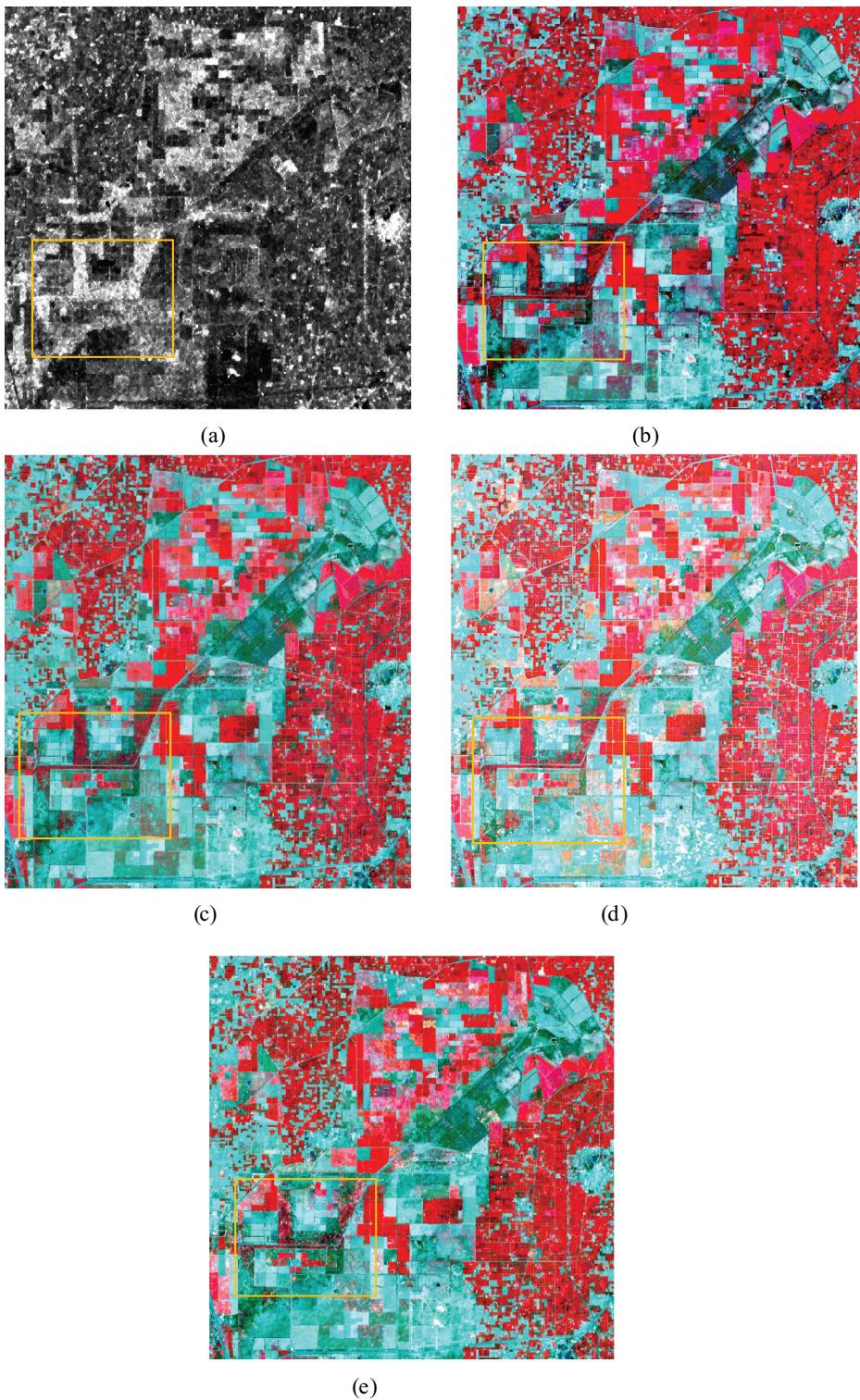


Figure 8. (a) S1 (VV), image fusion using (b) ROLP-IVF, (c) IVF, (d) MEDF and (e) MEF approaches. Yellow rectangle in figure represents the enhanced portion after fusion.

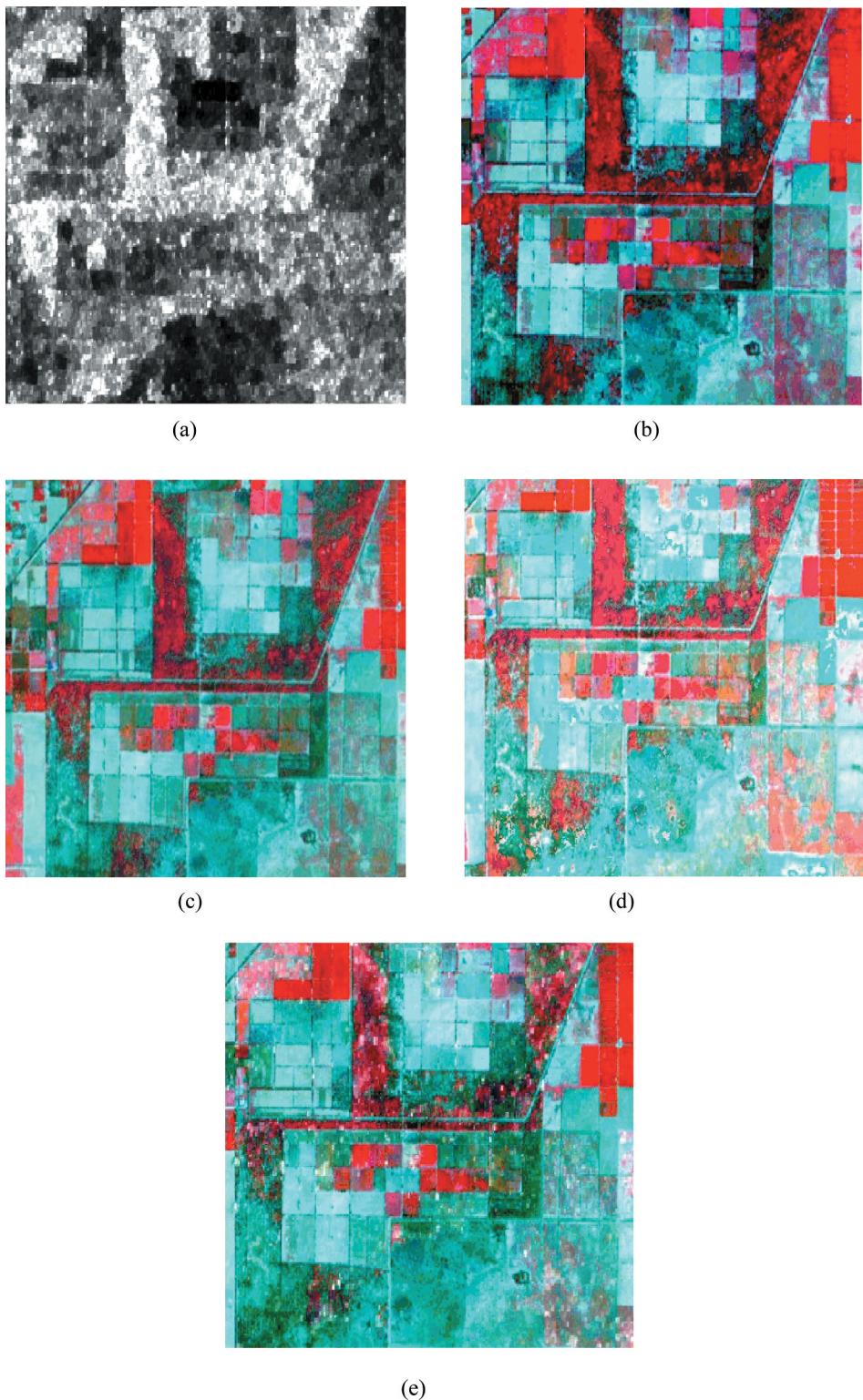


Figure 9. Zoomed Figure 8(a) S1 (VV), image fusion using (b) ROLP-IVF, (c) IVF, (d) MEDF and (e) MEF approaches.

5.2. Classification

To evaluate the performance of the fused images in terms of classification accuracy, classifications were carried out using a 2D-CNN classifier. Various images used for classification include: S2 data, S2 data layer stacked with S1 (VV) and S1 (VH), respectively, and fused images obtained by different fusion approaches. Overall classification accuracy and kappa values and class-wise accuracy for all datasets are provided in [Tables 2](#) and [3](#), respectively.

Results from [Table 2](#) suggest an improved performance by ROLP-IVF, IVF, MEF and MEDF-based fusion approaches in comparison to S2 data, S2 stacked with S1 (VV and VH separately) in terms of classification accuracy. Comparison of classification accuracy achieved by fused images obtained using different approaches suggests that the images obtained by ROLP-IVF approach achieved a significant gain in accuracy (~5%) using both VV and VH polarized S1 data in comparison to layer stacked S2 data with S1 (VV and VH) bands. Results from [Table 2](#) also suggest better performance by the fusion of S2 data with S1 (VV) band in comparison to S1 (VH) band by both ROLP-IVF and IVF approach. In comparison to the ROLP-IVF and IVF-based approach, MEF and MEDF approach works well with the fusion of S2 and S1 (VH) data, suggesting the importance of both VV and VH polarized S1 images for image fusion using S1 and S2 datasets.

[Table 3](#) provides class-wise accuracy by different fusion approaches. Results from the table suggest that the several land cover classes were classified with higher accuracy by the fused images obtained using ROLP-IVF, IVF, MEDF, and MEDF approaches. Based on improved performance by ROLP-IVF approach ([Table 2](#)), the class-wise accuracy obtained by this approach ([Table 3](#)) was compared with S2 data. Results from [Table 3](#) indicate an increase of 0% to 26% for different classes by ROLP-IVF approach-based fused images (S2 with S1 (VV) & S1 (VH)). On the other hand, comparison of class-wise accuracy achieved by ROLP-IVF-based fused image with the layer stacked (S2 with S1 (VV)) data suggest changes of the order of 0% to 21% in classification accuracy. In comparison to layer stacked image consisting of S2 and S1 (VH), an improvement of 0% to 22% in classification accuracy was achieved by ROLP-IVF approach for different classes.

Keeping in view of better performance by ROLP-IVF and IVF-based fused images (using S2 and S1 (VV)), classified images obtained by these two approaches as well as those by S2 only and S2 plus S1 (VV and VH layer staked) are provided in [Figure 10](#). To compare the change in classified images, a small part shown by a rectangle in [Figure 10](#) is zoomed and provided in [Figure 11](#).

Comparison of classified images suggests that S2 ([Figures 10–11\(a\)](#)), S2 with S1 (VH) ([Figures 10–11\(b\)](#)) and S1 (VV) ([Figures 10–11\(c\)](#)) have mixed classes as shown by the rectangular box in comparison to IVF and ROLP-IVF approaches ([Figures 10–11\(d–e\)](#)). Thus, it can clearly be seen that ROLP-IVF and IVF approaches ([Figures 10–11\(d–e\)](#)) works well in

Table 2. Classification results of S2, S2 with S1 (VV and VH separately), IVF, MEDF, MEF, and ROLP-IVF-based fusion approaches.

Accuracy measure	Data combination											
	S2+S1			IVF		MEF		MEDF		ROLP-IVF		
	S2	VV	VH	VV	VH	VV	VH	VV	VH	VV	VH	
OA (%)	89.18	91.94	91.25	95.29	92.99	91.00	93.45	90.23	93.41	96.11	95.16	
K	0.87	0.90	0.91	0.94	0.91	0.89	0.92	0.88	0.91	0.96	0.94	



Table 3. Class accuracy (%) with different datasets and methods: S2, S2 with S1 (VV and VH separately), IVF, MEF, MEDF, and ROLP-IVF-based fusion approaches.

Class	S2	Class accuracy (%)									
		S2+S1		IVF		MEF		MEDF		ROLP-IVF	
		VV	VH	VV	VH	VV	VH	VV	VH	VV	VH
1	91.00	95.00	98.00	96.00	93.00	98.00	100.00	97.00	98.00	99.00	98.00
2	92.00	99.00	98.00	98.00	97.00	99.00	93.00	97.00	89.00	99.00	98.00
3	97.00	96.00	99.00	92.00	89.00	97.00	98.00	98.00	97.00	99.00	99.00
4	74.00	81.00	80.00	64.00	74.00	81.00	71.00	81.00	95.00	89.00	88.00
5	100.00	86.00	96.00	89.00	96.00	86.00	86.00	82.00	79.00	89.00	64.00
6	85.00	79.00	89.00	94.00	85.00	85.00	88.00	80.00	85.00	93.00	92.00
7	93.00	97.00	98.00	98.00	97.00	93.00	97.00	99.00	98.00	99.00	99.00
8	81.00	87.00	77.00	87.00	82.00	79.00	88.00	66.00	88.00	92.00	89.00
9	95.00	85.00	80.00	96.00	86.00	82.00	83.00	73.00	79.00	89.00	87.00
10	84.00	97.00	91.00	94.00	93.00	93.00	88.00	88.00	93.00	89.00	95.00
11	73.00	88.00	90.00	92.00	90.00	88.00	82.00	88.00	79.00	93.00	93.00
12	88.00	78.00	83.00	97.00	83.00	83.00	85.00	85.00	81.00	88.00	90.00

terms of visual interpretation of the classified images also. A possible reason of improved performance by ROLP-IVF approach is that this approach was able to preserve the shape of the vegetation leading to better classification results.

6. Conclusion

This study reports the results of CNN-based fusion approaches using S1 (VV and VH separately) and S2 data. Based on the performance of various fusion methods, one of the primary conclusions is that out of three existing fusion approaches used in the study, only IVF approach performed well, suggesting the data specific nature of CNN-based fusion approaches. Another major conclusion drawn from this study is that the modified IVF approach (ROLP-IVF) as proposed in this study outperformed existing IVF approach in terms of various indicators. Though the use of VV and VH polarization of S1 image was used for image fusion, results suggest better performance by VV polarized S1 image in comparison to VH polarized image in terms of classification accuracy. Thus, suggesting a need for further studies using different study areas consisting of different land cover classes to evaluate the effect of polarization on classification accuracy during image fusion.

Keeping in view of the varied performances by different fusion approaches, future study is also planned to implement the various pyramidal methods with the remaining fusion approaches and utilizing these approaches for multi-temporal dataset over the study area.

Acknowledgements

Authors are thankful to Space and Application Centre (SAC), Ahmedabad, for the financial support under the project titled "Fusion of Optical and Multi-frequency Multi-polarimetric SAR data for Enhanced Land cover Mapping" to carry out this work. The authors would like to acknowledge the National Institute of Technology, Kurukshetra, India for providing the required computing facilities to carry out this study.

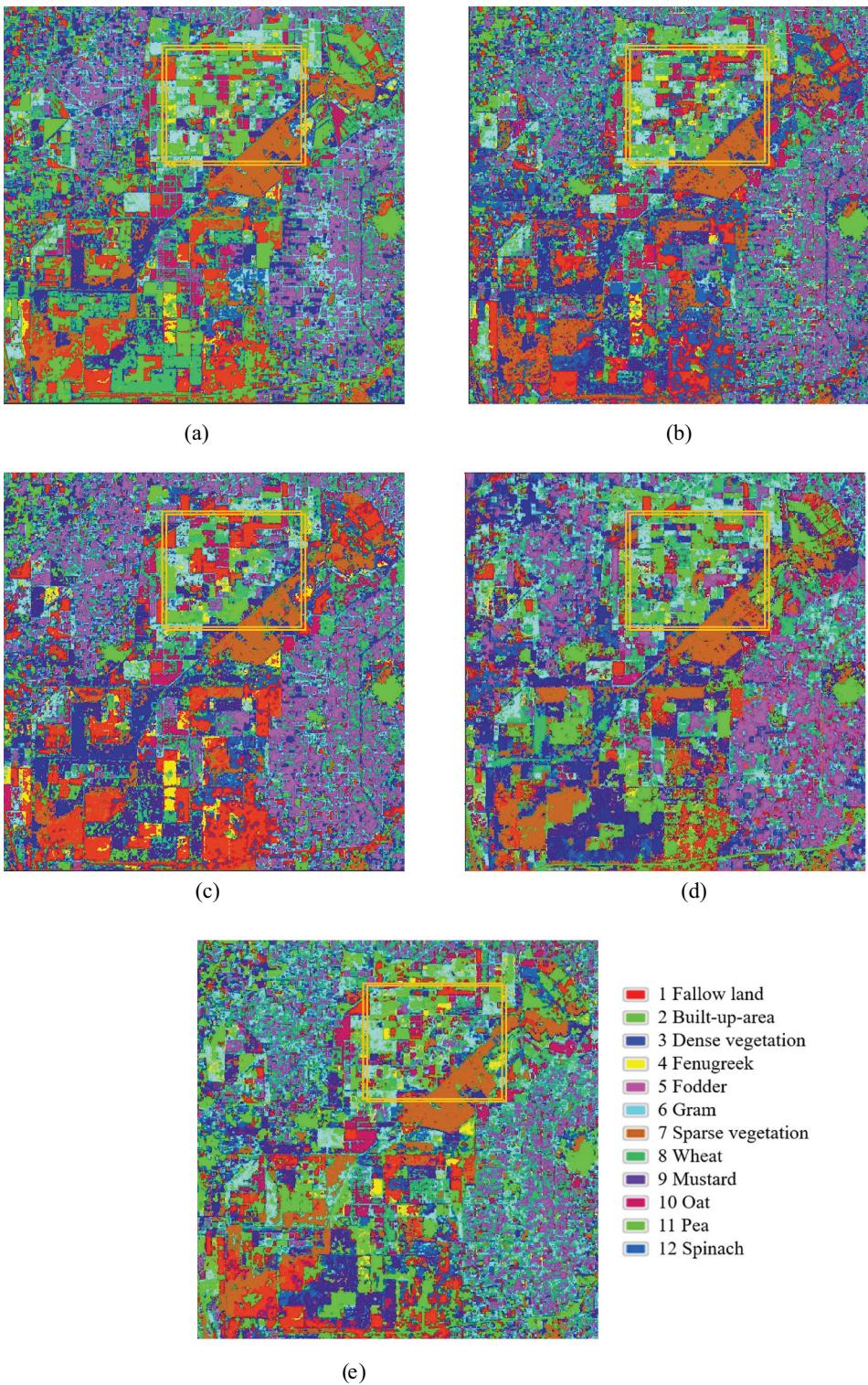


Figure 10. Classified images of (a) S2 image, (b) S2+S1 (VH), (c) S2+S1 (VV), (d) IVF (VV) and (e) ROLP-IVF (VV) approaches using 2D-CNN classifier. Yellow rectangle in figure represents the enhanced portion after fusion.

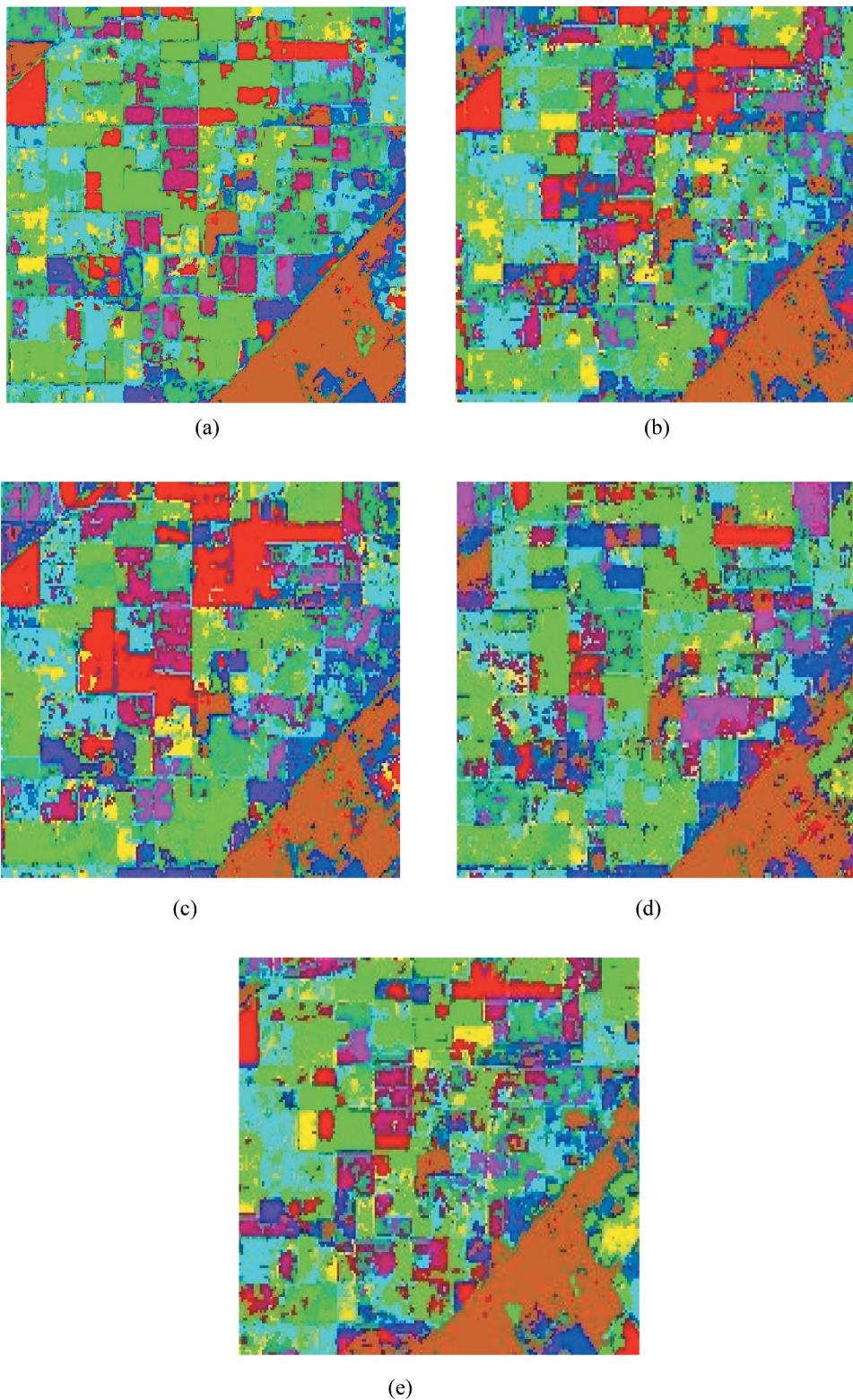


Figure 11. Zoomed classified images of Figure 10(a) S2 image, (b) S2+S1 (VH), (c) S2+S1 (VV), (d) IVF (VV) and (e) ROLP-IVF (VV) approaches using 2D-CNN classifier. Yellow rectangle in figure represents the enhanced portion after fusion.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Achala Shakya  <http://orcid.org/0000-0003-2880-5743>
 Mantosh Biswas  <http://orcid.org/0000-0001-9027-4432>
 Mahesh Pal  <http://orcid.org/0000-0003-1805-2952>

References

- Abdikan, S., S. Fusun, S. Filiz, and E. Manfred. 2014. "A Comparative Data-fusion Analysis of Multi-sensor Satellite Images." *International Journal of Digital Earth* 7 (8): 671–687. doi:[10.1080/17538947.2012.748846](https://doi.org/10.1080/17538947.2012.748846).
- Aiazzi, B., S. Baronti, and M. Selva. 2007. "Improving Component Substitution Pansharpening through Multivariate Regression of MS + Pan Data." *IEEE Transactions on Geoscience and Remote Sensing* 45 (10): 3230–3239. doi:[10.1109/TGRS.2007.901007](https://doi.org/10.1109/TGRS.2007.901007).
- Al-Ahmadi, F. S., and A. S. Hames. 2009. "Comparison of Four Classification Methods to Extract Land Use and Land Cover from Raw Satellite Images for Some Remote Arid Areas." *Kingdom of Saudi Arabia, Earth* 20: 167–191.
- Aschbacher, J., and J. Lichtenegger. 1990. "Complementary Nature of SAR and Optical Data: A Case Study in the Tropics." *Earth Observation Quarterly* 31: 4–8.
- Azarang, A., and H. Ghassemian. 2018. "Application of Fractional-order Differentiation in Multispectral Image Fusion." *Remote Sensing Letters* 9 (1): 91–100. doi:[10.1080/2150704X.2017.1395963](https://doi.org/10.1080/2150704X.2017.1395963).
- Benedetti, P., D. Ienco, R. Gaetano, K. Ose, R. G. Pensa, and S. Dupuy. 2018. "M3 Fusion: A Deep Learning Architecture for Multiscale Multimodal Multitemporal Satellite Data Fusion." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11 (12): 4939–4949. doi:[10.1109/JSTARS.2018.2876357](https://doi.org/10.1109/JSTARS.2018.2876357).
- Borzov, S.M., Potaturkin, O.I. 2018. "Classification of Hyperspectral Images with Different Methods of Training Set Formation." *Optoelectronics, Instrumentation and Data Processing* 54: 76–82. <https://doi.org/10.3103/S8756699018010120>.
- Burt, P., and E. Adelson. 1983. "The Laplacian Pyramid as a Compact Image Code." *IEEE Transactions on Communications* 31 (4): 532–540. doi:[10.1109/TCOM.1983.1095851](https://doi.org/10.1109/TCOM.1983.1095851).
- Clerici, N., C. A. Valbuena Calderón, and J. M. Posada. 2017. "Fusion of Sentinel-1A and Sentinel-2A Data for Land Cover Mapping: A Case Study in the Lower Magdalena Region, Colombia." *Journal of Maps* 13 (2): 718–726. doi:[10.1080/17445647.2017.1372316](https://doi.org/10.1080/17445647.2017.1372316).
- Das, M., and S. K. Ghosh. 2016. "Deep-STEP: A Deep Learning Approach for Spatiotemporal Prediction of Remote Sensing Data." *IEEE Geoscience and Remote Sensing Letters* 13 (12): 1984–1988. doi:[10.1109/LGRS.2016.2619984](https://doi.org/10.1109/LGRS.2016.2619984).
- Ghamisi, P., E. Maggiori, S. Li, R. Souza, Y. Tarablaka, G. Moser, and S. B. Serpico. 2018. "New Frontiers in Spectral-spatial Hyperspectral Image Classification: The Latest Advances Based on Mathematical Morphology, Markov Random Fields, Segmentation, Sparse Representation, and Deep Learning." *IEEE Geoscience and Remote Sensing Magazine* 6 (3): 10–43. doi:[10.1109/MGRS.2018.2854840](https://doi.org/10.1109/MGRS.2018.2854840).
- Ghassemian, H., 2016. "A review of remote sensing image fusion methods." *Information Fusion* 32: 75–89.
- Ghassemian, H. 2016. "A Review of Remote Sensing Image Fusion Methods." *Information Fusion* 32: 75–89. doi:[10.1016/j.inffus.2016.03.003](https://doi.org/10.1016/j.inffus.2016.03.003).
- Gibril, M., A. Barakat, S. A. Bakar, K. Yao, M. O. Idrees, and B. Pradhan. 2017. "Fusion of RADARSAT-2 and Multispectral Optical Remote Sensing Data for LULC Extraction in a Tropical Agricultural Area." *Geocarto International* 32 (7): 735–748. doi:[10.1080/10106049.2016.1170893](https://doi.org/10.1080/10106049.2016.1170893).
- Glorot, X., A. Bordes, and Y. Bengio. 2011. "Deep Sparse Rectifier Networks." International conference on artificial intelligence and statistics, 315–323. Fort Lauderdale, FL, USA.



- Goodfellow, I., Y. Bengio, and A. Courville. 2016. *Deep Learning*. Cambridge, Massachusetts: MIT Press. [Online]. <http://www.deeplearningbook.org>
- Gulcehre, C., K. Cho, R. Pascanu, and Y. Bengio. 2014. "Learned-norm Pooling for Deep Feedforward and Recurrent Neural Networks." Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 530–546, Berlin, Heidelberg.
- He, H., M. Chen, T. Chen, D. Li, and P. Cheng. 2019. "Learning to Match Multitemporal Optical Satellite Images Using Multi-support-patches Siamese Networks." *Remote Sensing Letters* 10 (6): 516–525. doi:[10.1080/2150704X.2019.1577572](https://doi.org/10.1080/2150704X.2019.1577572).
- He, W. and Yokoya, N., 2018. Multi-Temporal Sentinel-1 and -2 Data Fusion for Optical Image Simulation." *ISPRS International Journal of Geo-Information* 7(10): 389–400 doi:[10.3390/ijgi7100389](https://doi.org/10.3390/ijgi7100389).
- Hughes, L. H., M. Schmitt, L. Mou, Y. Wang, and X. X. Zhu. 2018. "Identifying Corresponding Patches in SAR and Optical Images with a Pseudo-siamese CNN." *Geoscience and Remote Sensing Letters* 15 (5): 784–788. doi:[10.1109/LGRS.2018.2799232](https://doi.org/10.1109/LGRS.2018.2799232).
- Hughes, L. H., N. Merkle, T. Bürgmann, S. Auer, and M. Schmitt. 2019. "Deep Learning for SAR-Optical Image Matching." *IEEE International Geoscience and Remote Sensing* 4877–4880, Yokohama, Japan..
- Hughes, L.H., Merkle, N., Bürgmann, T., Auer, S. and Schmitt, M., 2019. "Deep Learning for SAR-Optical Image Matching." IEEE International Geoscience and Remote Sensing Symposium, 4877–4880. Yokohama, Japan
- Jahan, F., and M. Awrangjeb. 2017. "Pixel-based Land Cover Classification by Fusing Hyperspectral and LiDAR Data. International Archives of the Photogrammetry." *Remote Sensing and Spatial Information Sciences* 42 (27): 711–718.
- Jia, Y., C. Huang, and T. Darrell. 2012. "Beyond Spatial Pyramids: Receptive Field Learning for Pooled Image Features." IEEE Conference on Computer Vision and Pattern Recognition, 3370–3377, Providence, RI, USA.
- Kamilaris, A., and F. X. Prenafeta-Boldú. 2018. "Deep Learning in Agriculture: A Survey." *Computers and Electronics in Agriculture* 147: 70–90. doi:[10.1016/j.compag.2018.02.016](https://doi.org/10.1016/j.compag.2018.02.016).
- Karathanassi, V., P. Kolokousis, and S. Ioannidou. 2007. "A Comparison Study on Fusion Methods Using Evaluation Indicators." *International Journal of Remote Sensing* 28 (10): 2309–2341. doi:[10.1080/01431160600606890](https://doi.org/10.1080/01431160600606890).
- Kim, Y. 2014. *Convolutional Neural Networks for Sentence Classification*. arXiv Preprint. arXiv:1408.5882.
- Kingma, D. P., and J. Ba. 2014. "Adam: A Method for Stochastic Optimization." arXiv preprint arXiv:1412.6980. Accessed on 15-01-2020.
- Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. "Imagenet Classification with Deep Convolutional Neural Networks." Conference on Neural Information Processing Systems, 1097–1105, Lake Tahoe, United States.
- Kulkarni, S. C., and P. P. Rege. 2020. "Pixel Level Fusion Techniques for SAR and Optical Images: A Review." *Information Fusion* 59: 13–29. doi:[10.1016/j.inffus.2020.01.003](https://doi.org/10.1016/j.inffus.2020.01.003).
- Laben, C. A., B. V. Brower, and E. K. Co. 2000. "Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-sharpening." <https://www.google.com/patents/US6011875>
- Lal, A. M., and S. M. Anouncia. 2016. "Enhanced Dictionary Based Sparse Representation Fusion for Multi-temporal Remote Sensing Images." *European Journal of Remote Sensing* 49 (1): 317–336. doi:[10.5721/EJRS20164918](https://doi.org/10.5721/EJRS20164918).
- LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner. 1998. "Gradient-based Learning Applied to Document Recognition." *Proceedings of the IEEE* 86 (11): 2278–2324. doi:[10.1109/5.726791](https://doi.org/10.1109/5.726791).
- Lee, J., D. Han, M. Shin, J. Im, J. Lee, and L. J. Quackenbush. 2020. "Different Spectral Domain Transformation for Land Cover Classification Using Convolutional Neural Networks with Multi-Temporal Satellite Imagery." *Remote Sensing* 12 (7): 1097–1125. doi:[10.3390/rs12071097](https://doi.org/10.3390/rs12071097).
- Li, W., R. Dong, H. Fu, J. Wang, L. Yu, and P. Gong. 2020. "Integrating Google Earth Imagery with Landsat Data to Improve 30-m Resolution Land Cover Mapping." *Remote Sensing of Environment* 237: 111–563. doi:[10.1016/j.rse.2019.111563](https://doi.org/10.1016/j.rse.2019.111563).
- LiKamWa, R., Y. Hou, J. Gao, M. Polansky, and L. Zhong. 2016. "RedEye: Analog ConvNet Image Sensor Architecture for Continuous Mobile Vision." *ACM SIGARCH Computer Architecture News* 44 (3): 255–266. doi:[10.1145/3007787.3001164](https://doi.org/10.1145/3007787.3001164).

- Lillesand, T. M., R. W. Kiefer, and J. W. Chipman. 2015. *Remote Sensing and Image Interpretation*. New York: John Wiley & Sons.
- Liu, J. G. 2000. "Smoothing Filter-based Intensity Modulation: A Spectral Preserve Image Fusion Technique for Improving Spatial Details." *International Journal of Remote Sensing* 21 (18): 3461–3472. doi:[10.1080/014311600750037499](https://doi.org/10.1080/014311600750037499).
- Liu, Y., X. Chen, J. Cheng, and H. Peng. 2017a. "A Medical Image (MEDF) Fusion Method Based on Convolutional Neural Networks." *International Conference on Information Fusion* 20: 1–7.
- Liu, Y., X. Chen, J. Cheng, H. Peng, and Z. Wang. 2017b. "Multi-focus Image (MEF) Fusion with a Deep Convolutional Neural Network." *Information Fusion* 36: 191–207. doi:[10.1016/j.inffus.2016.12.001](https://doi.org/10.1016/j.inffus.2016.12.001).
- Liu, Y., X. Chen, J. Cheng, H. Peng, and Z. Wang. 2018. "Infrared and Visible Image (IVF) Fusion with Convolutional Neural Networks. International Journal of Wavelets." *Multiresolution and Information Processing* 16 (3): 1850018–1850037. doi:[10.1142/S0219691318500182](https://doi.org/10.1142/S0219691318500182).
- Ma, J., Y. Ma, and C. Li. 2019. "Infrared and Visible Image Fusion Methods and Applications: A Survey." *Information Fusion* 45: 153–178. doi:[10.1016/j.inffus.2018.02.004](https://doi.org/10.1016/j.inffus.2018.02.004).
- Maggiori, E., Y. Tarabalka, G. Charpiat, and P. Alliez. 2016. "Convolutional Neural Networks for Large-scale Remote-sensing Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 55 (2): 645–657. doi:[10.1109/TGRS.2016.2612821](https://doi.org/10.1109/TGRS.2016.2612821).
- Márquez, C., López, M.I., Ruisánchez, I. and Callao, M.P., 2016. "FT-Raman and NIR spectroscopy data fusion strategy for multivariate qualitative analysis of food fraud." *Talanta* 161: 80–86. doi:[10.1016/j.talanta.2016.08.003](https://doi.org/10.1016/j.talanta.2016.08.003).
- Masi, G., D. Cozzolino, L. Verdoliva, and G. Scarpa. 2016. "Pansharpening by Convolutional Neural Networks." *Remote Sensing* 8 (7): 1–22. doi:[10.3390/rs8070594](https://doi.org/10.3390/rs8070594).
- Mazzia, V., A. Khalil, and M. Chiaberge. 2020. "Improvement in Land Cover and Crop Classification Based on Temporal Features Learning from Sentinel-2 Data Using Recurrent-Convolutional Neural Network (R-CNN)." *Applied Sciences* 10 (1): 238–261. doi:[10.3390/app10010238](https://doi.org/10.3390/app10010238).
- Meng, X., H. Shen, L. Huifang, L. Zhang, and F. Randi. 2019. "Review of the Pansharpening Methods for Remote Sensing Images Based on the Idea of Meta-analysis: Practical Discussion and Challenges." *Information Fusion* 46: 102–113. doi:[10.1016/j.inffus.2018.05.006](https://doi.org/10.1016/j.inffus.2018.05.006).
- Mountrakis, G., J. Im, and C. Ogole. 2011. "Support Vector Machines in Remote Sensing: A Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (3): 247–259. doi:[10.1016/j.isprsjprs.2010.11.001](https://doi.org/10.1016/j.isprsjprs.2010.11.001).
- Olkkinen, H., and P. Pesola. 1996. "Gaussian Pyramid Wavelet Transform for Multiresolution Analysis of Images." *Graphical Models and Image Processing* 58 (4): 394–398. doi:[10.1006/gmip.1996.0032](https://doi.org/10.1006/gmip.1996.0032).
- Palsson, F., J. R. Sveinsson, and M. O. Ulfarsson. 2017. "Multispectral and Hyperspectral Image Fusion Using a 3-D-Convolutional Neural Network." *IEEE Geoscience and Remote Sensing Letters* 14 (5): 639–643. doi:[10.1109/LGRS.2017.2668299](https://doi.org/10.1109/LGRS.2017.2668299).
- Parihar, N., V. S. Rathore, and S. Mohan. 2017. "Combining ALOS PALSAR and AVNIR-2 Data for Effective Land Use/land Cover Classification in Jharia Coalfields Region." *International Journal of Image and Data Fusion* 8 (2): 130–147. doi:[10.1080/19479832.2016.1273258](https://doi.org/10.1080/19479832.2016.1273258).
- Petropoulos, G. P., K. P. Vadrevu, and C. Kalaitzidis. 2013. "Spectral Angle Mapper and Object-based Classification Combined with Hyperspectral Remote Sensing Imagery for Obtaining Land Use/cover Mapping in a Mediterranean Region." *Geocarto International* 28 (2): 114–129. doi:[10.1080/10106049.2012.668950](https://doi.org/10.1080/10106049.2012.668950).
- Pohl, C., and J. L. Van Genderen. 1998. "Multisensor Image Fusion in Remote Sensing: Concepts, Methods and Applications." *Remote Sensing* 5 (5): 823–854. doi:[10.1080/014311698215748](https://doi.org/10.1080/014311698215748).
- Reiche, J., S. De Bruin, D. Hoekman, J. Verbesselt, and M. Herold. 2015. "A Bayesian Approach to Combine Landsat and ALOS PALSAR Time Series for near Real-time Deforestation Detection." *Remote Sensing* 7 (5): 4973–4996. doi:[10.3390/rs70504973](https://doi.org/10.3390/rs70504973).
- Rogers, R., and L. Wood. 1990. "The History and Status of Merging Multiple Sensor Data- an Overview." *ACSM-ASPRS Annual Conference Image Processing and Remote Sensing* 4: 352–360.
- Scarpa, G., M. Gargiulo, A. Mazza, and R. Gaetano. 2018. "A CNN-based Fusion Method for Feature Extraction from Sentinel Data." *Remote Sensing* 10 (2): 1–20. doi:[10.3390/rs10020236](https://doi.org/10.3390/rs10020236).
- Schmitt, M., Tupin, F. and Zhu, X.X., 2017. "Fusion of SAR and optical remote sensing data—Challenges and recent trends." International Geoscience and Remote Sensing Symposium (IGARSS), 5458–5461. Fort Worth, TX, USA.

- Schmitt, M., and X. X. Zhu. 2016. "Data Fusion and Remote Sensing – An Ever-growing Relationship." *IEEE Geoscience and Remote Sensing Magazine* 4 (4): 6–23. doi:[10.1109/MGRS.2016.2561021](https://doi.org/10.1109/MGRS.2016.2561021).
- Sheoran, A., and B. Haack. 2014. "Optical and Radar Data Comparison and Integration: Kenya Example." *Geocarto International* 29 (4): 370–382. doi:[10.1080/10106049.2013.769027](https://doi.org/10.1080/10106049.2013.769027).
- Toet, A. 1989. "Image Fusion by a Ratio of Low-pass Pyramid." *Pattern Recognition Letters* 9 (4): 245–253. doi:[10.1016/0167-8655\(89\)90003-2](https://doi.org/10.1016/0167-8655(89)90003-2).
- Tran, D., J. Ray, Z. Shou, S. F. Chang, and M. Paluri. 2017. *Convnet Architecture Search for Spatiotemporal Feature Learning*, 1–12. arXiv preprint arXiv:1708.05038.
- Tripathi, G., A. C. Pandey, B. R. Parida, and A. Kumar. 2020. "Flood Inundation Mapping and Impact Assessment Using Multi-Temporal Optical and SAR Satellite Data: A Case Study of 2017 Flood in Darbhanga District, Bihar, India." *Water Resources Management* 34 (6): 1–22. doi:[10.1007/s11269-020-02534-3](https://doi.org/10.1007/s11269-020-02534-3).
- Tripathi, G., B. R. Parida, and A. C. Pandey. 2019. "Spatio-Temporal Rainfall Variability and Flood Prognosis Analysis Using Satellite Data over North Bihar during the August 2017 Flood Event." *Hydrology* 6 (2): 38–61. doi:[10.3390/hydrology6020038](https://doi.org/10.3390/hydrology6020038).
- Vivone, G. 2020. "Pansharpening." *Data Handling in Science and Technology* 32: 69–91.
- Vivone, G., L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald. 2015. "A Critical Comparison among Pansharpening Algorithms." *IEEE Transactions on Geoscience and Remote Sensing* 53 (5): 2565–2586. doi:[10.1109/TGRS.2014.2361734](https://doi.org/10.1109/TGRS.2014.2361734).
- Wang, W., and F. Chang. 2011. "A Multi-focus Image Fusion Method Based on Laplacian Pyramid." *Journal of Computers* 6 (12): 2559–2566. doi:[10.4304/jcp.6.12.2559-2566](https://doi.org/10.4304/jcp.6.12.2559-2566).
- Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. 2004. "Image Quality Assessment: From Error Measurement to Structural Similarity." *IEEE Transactions on Image Processing* 13 (1): 600–612. doi:[10.1109/TIP.2003.819861](https://doi.org/10.1109/TIP.2003.819861).
- Weng, Q., D. Quattrochi, and P. E. Gamba. 2018. *Urban Remote Sensing*. Boca Raton: CRC press/Taylor and Francis.
- Yu, S., S. Jia, and C. Xu. 2017. "Convolutional Neural Networks for Hyperspectral Image Classification." *Neurocomputing* 219: 88–98. doi:[10.1016/j.neucom.2016.09.010](https://doi.org/10.1016/j.neucom.2016.09.010).
- Zagoruyko, S., and N. Komodakis. 2015. "Learning to Compare Image Patches via Convolutional Neural Networks." *IEEE Conference on Computer Vision and Pattern Recognition*, 1: 4353–4361.
- Zhang, C., P. Yue, D. Tapete, B. Shangguan, M. Wang, and Z. Wu. 2020. "A Multi-level Context-guided Classification Method with Object-based Convolutional Neural Network for Land Cover Classification Using Very High Resolution Remote Sensing Images." *International Journal of Applied Earth Observation and Geoinformation* 88: 102086–102099. doi:[10.1016/j.jag.2020.102086](https://doi.org/10.1016/j.jag.2020.102086).
- Zhang, H., H. Lin, and Y. Li. 2015. "Impacts of Feature Normalization on Optical and SAR Data Fusion for Land Use/Land Cover Classification." *IEEE Geoscience and Remote Sensing Letters* 12 (5): 1061–1065. doi:[10.1109/LGRS.2014.2377722](https://doi.org/10.1109/LGRS.2014.2377722).
- Zhong, J., Yang, B., Huang, G., Zhong, F., & Chen, Z. 2016. "Remote Sensing Image Fusion with Convolutional Neural Network." *Sensing and Imaging* 17(1): 1–16. doi:[10.1007/s11220-016-0135-6](https://doi.org/10.1007/s11220-016-0135-6)
- Zhong, L., L. Hu, and H. Zhou. 2019. "Deep Learning Based Multi-temporal Crop Classification." *Remote Sensing of Environment* 221: 430–443. doi:[10.1016/j.rse.2018.11.032](https://doi.org/10.1016/j.rse.2018.11.032).
- Zhou, Z., B. Wang, S. Li, and M. Dong. 2016. "Perceptual Fusion of Infrared and Visible Images through a Hybrid Multi-scale Decomposition with Gaussian and Bilateral filters." *Information Fusion* 30: 15–26. doi:[10.1016/j.inffus.2015.11.003](https://doi.org/10.1016/j.inffus.2015.11.003).
- Zhou, Z., S. Li, and B. Wang. 2014. "Multi-scale Weighted Gradient-based Fusion for Multi-focus Images." *Information Fusion* 20: 60–72. doi:[10.1016/j.inffus.2013.11.005](https://doi.org/10.1016/j.inffus.2013.11.005).
- Zhu, X. X., C. Grohnfeldt, and R. Bamler. 2015. "Exploiting Joint Sparsity for Pansharpening: The J-SparseFI Algorithm." *IEEE Transactions on Geoscience and Remote Sensing* 54 (5): 2664–2681. doi:[10.1109/TGRS.2015.2504261](https://doi.org/10.1109/TGRS.2015.2504261).