DEVELOPMENT OF DEEP LEARNING APPROACHES TO EXTRACT INFORMATION FROM SATELLITE IMAGES

Ph.D. THESIS

by

GOPAL SINGH PHARTIYAL

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE, UTTARAKHAND -247 667 (INDIA)
JUNE, 2021

DEVELOPMENT OF DEEP LEARNING APPROACHES TO EXTRACT INFORMATION FROM SATELLITE IMAGES

A THESIS

Submitted in the partial fulfilment of the requirements for the award of the degree

of

DOCTOR OF PHILOSOPHY

in

ELECTRONICS AND COMMUNICATION ENGINEERING

by

GOPAL SINGH PHARTIYAL

**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE, UTTARAKHAND -247 667 (INDIA)
JUNE, 2021**

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

STUDENT'S DECLARATION

I hereby certify that the work presented in the thesis entitled "**DEVELOPMENT OF DEEP LEARNING APPROACHES TO EXTRACT INFORMATION FROM SATELLITE IMAGES**" in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy and submitted in the Department of Electronics And Communication Engineering of Indian Institute of Technology Roorkee, Roorkee is my own work carried out during a period from **July, 2015** to **June, 2021** under the supervision of **Dr Dharmendra Singh,** Professor, Department of Electronics and Communication Engineering of Indian Institute of Technology Roorkee, Roorkee, India and **Dr Hussein Yahia,** Professor, Head, project team "GeoStat", INRIA (Institut National de Recherche en Informatique et en Automatique), Bordeaux, France**.**

The matter presented in this thesis has not been submitted for the award of any other degree of this or any other Institute.

Dated: June 15, 2021   **Gopal Singh Phartiyal**

SUPERVISOR'S DECLARATION

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Dr Dharmendra Singh   **Dr Hussein Yahia**
(Supervisor)   (Supervisor)

Dated: June 15, 2021

To

My Parents, God and, Friends

ABSTRACT

This thesis titled 'Development of Deep Learning Approaches to Extract Information from Satellite Images' incorporates the study and critical analysis of the performance of conventional convolutional neural network (CNN) and recurrent neural network (RNN) models, and further development of novel CNN and RNN models to exploit the spectral and/or spatial and/or temporal information present in satellite images for applications such as land cover classification, disaster monitoring and mitigation, and agriculture crop monitoring. The main objective of this thesis is to pursue these studies with unimodal or multimodal satellite images. Special focus during these studies is to develop novel deep neural network (DNN) models (common term for CNN and RNN based models) that can achieve better classification accuracy and generalization performance in mixed land cover class scenarios.

Four tasks are set to pursue these studies. The tasks set are based on: type of satellite image used, application for which the task is realized, multi-modality images involvement, and presence of time series information. The tasks to be carried out are; i) Study and assessment of polarization signatures as potential features for PolSAR image based land cover classification and further development of novel polarization signatures and CNNs based approaches to achieve improved classification performances with PolSAR satellite images, ii) Development of a novel approach using polarization signatures and CNNs for detection of human settlements affected during floods, iii) Development of novel spectral and spectral-spatial CNN models for multisensor (PolSAR and multispectral) images based land cover classification, and iv) Development of novel CNN, RNN, and CNN-RNN models for multisensor (PolSAR and multispectral) satellite time series images based crop classification.

The satellite image modalities used in these studies are PolSAR and multispectral images because these two satellite image modalities are most common and used popularly. Therefore, development of novel DNN models on these modalities is motivating and beneficial. ALOS PALSAR-2 and Sentinel-1 satellite PolSAR images and Landsat-8 and Sentinel-2 satellite multispectral images are considered in these studies. Since, SAR images are all weather images and are recently the popular satellite image modality, special emphasis is given on study and development of CNN and RNN models to extract information from PolSAR images, in standalone or in combination with multispectral images. Multiple study areas are also considered for testing of the developed models or approaches. The study areas are selected based on presence of diversity in land cover (mixed land cover) and ease of accessibility to ground truth data. Extensive ground truth collection is carried out in the study areas for the development and testing

of the novel models or approaches. Thousands of ground truth samples are collected for each task throughout the study period. The ground truth samples are either collected directly on the terrain using GPS or via visual inspection of the satellite images used in different studies and corresponding Google Earth images.

The thesis consists of seven chapters. ***Chapter 1*** provides; a brief introduction to satellite images, major challenges involved with traditional satellite image processing techniques, motivation to develop deep learning and DNN based models for satellite image processing, major objectives to be carried out during the research, and framework & organisation of the thesis.

Chapter 2 provides a brief literature review on the several state-of-the art deep learning and DNN based models and techniques developed for various satellite image based applications. The review includes literature on DNN based techniques developed for single senor/unimodal satellite images, multisensor/multimodal satellite images, and satellite time series images.

The study presented in ***Chapter 3*** involves a critical analysis of polarization signatures (PSs) as potential features for PolSAR data based land cover classification and development of novel approaches which utilize polarization signatures to provide improved classification performance. The study is divided into three sub tasks i.e. *sub task 1:* critical analysis of PSs as potential PolSAR features, *sub task 2*: development of an optimal class boundary estimator algorithm to fully exploit PSs, and *sub task 3*: development of CNN model to exploit PSs. The first sub task of this study explores PSs and utilizes it for the computation of PSs based features from quad PolSAR data. A land cover (e.g. bare soil) behaves as a complex mixture of different canonical scatterers or canonical targets and therefore its PS is difficult to interpret. Therefore, there is a need to find correlations between PSs of land covers and PSs of canonical targets. These correlations indicate the degree of similarity between the two scatterers. In the direction, polarization signature correlation features (PSCFs) are obtained from PSs. Further, the PSCFs are critically analysed as potential features for PolSAR image based LC classification. The analysis indicated to the significant potential of polarization signatures or PSCFs as features for PolSAR image based land cover classification. In the second and third sub tasks of the study, two different approaches are developed. In particular, both classification techniques are developed to improve classification between tall vegetation and built-up areas because the built-up areas also produce volume scattering response along with double bounce which makes segregation of the two classes challenging in a PolSAR image. In the second sub task of the study, i.e. *sub task 2*, a novel adaptive and optimal class boundary estimator based decision tree classification approach is developed. Image statistics are used to obtain class boundary values

which make the approach robust and adaptive. Genetic algorithms are used to optimize the class boundary values to achieve maximum overall classification accuracy. Finally, a decision tree (DT) classifier, which utilizes the optimal class boundaries obtained by using this novel estimator on PSCFs, is realized. At last, in the third sub task of the study, i.e. *sub task 3*, a second novel classification approach namely PSs-CNN which utilizes 2D-CNNs as classifier directly with polarization signatures to perform land cover classification is developed. Since, the PSs are two dimensional representation of target backscatter response, they can be treated as images of target response and thus can intuitively be used in 2D-CNN classification framework. The CNNs captures the spatial context present in a pixel PSs which helps in its classification. For experimentation, PALSAR-2 fully polarimetric SAR data over Roorkee (Uttarakhand, India) study area is considered. Five land cover classes are considered namely built-up, dense tall vegetation such as forests, bare soil, short vegetation such as agriculture crops, and water. Both novel approaches are compared with other popular methods. The optimal class boundary estimator based DT classifier reports highest classification accuracy of 75.3% whereas the PSs-CNN classifier reports 73.2.8% classification accuracy. Classification results from the novel PSs-CNN model are comparable to the optimal class boundary estimator based DT classifier. However, the CNN technique is better than the optimal class boundary estimator based DT classifier because the PSs-CNN classifier is simple, less hand-crafted, and easily scalable.

Satellite PolSAR images are the most reliable, useful, and consistent form of land cover information during a flood disaster because of its weather independent monitoring capabilities. *Chapter 4* presents an efficient approach to map human settlements, especially villages in the rural parts of India that are affected during flood using PolSAR imagery, CNNs, *apriori* knowledge, and image morphology. Monitoring human settlements in the rural parts of India during floods is difficult even with PolSAR data. The built-up structures (e.g. houses) in these human settlements, especially villages of rural India are closely packed and randomly oriented. Therefore, these villages show volume scattering response along with double bounce scattering response in a PolSAR image. The mixed scattering response from these villages hinders their segregation from tall vegetation in a PolSAR image. This challenge is also discussed in chapter 3. Therefore, the detection and mapping of these villages during a flood using PolSAR image is challenging as there is high risk of false alarms, or missing out a flood affected village during rescue operations.

The approach involves three stages: –*first*, built-up area extraction from a PolSAR image, –*second*, detection of villages in a built-up area image, and –*third*, identification and mapping villages that are affected by flood and masking out villages that are not. In the first stage, a CNN

based novel classifier which utilizes PolSAR polarization signatures (PSs) as features (similar to as in chapter 3), is proposed to extract built-up areas from a PolSAR image. This novel classifier uses a window (rather than a pixel) of PSs as input image because the villages are usually large in area and therefore cover several pixels in a PolSAR image. The window approach helps suppress the volume scattering present in the backscatter response of the villages and thereby improves the detection of these villages. However, there is presence of false alarms and true negatives in the built-up areas extracted and there is a need to mask out these artefacts. Therefore, the second stage proposes a novel village detection methodology. The methodology detects whether a particular object in the built-up area image (obtained in the first stage) is a village or not. A series of morphological operations are developed based on *apriori* knowledge are employed to suppress false alarms and true negatives. Next, villages that are affected by flood are required to be mapped and villages that are not affected by flood are masked out. In the third stage, flood affected villages are identified and mapped and non-flooded villages are masked out using another image morphology based methodology.

Quantitative and qualitative assessment of the performance of the novel approach over two different study areas is carried out. Different artificial flood scenarios are created to evaluate the performance of developed approach. The developed approach is evaluated under these flood scenarios and flood affected villages are identified and mapped. Finally, number of villages successfully detected and false alarms are computed with the help of ground truth information (actual number of villages or towns in the study regions are known). Results indicate to high detection accuracy (above 90%), low false alarms and true negatives. The approach is scalable and potentially satisfactory for real time usage.

Chapter 5 presents study and analysis of conventional, and further development of novel multisensor satellite image based land cover classification approaches that are based on CNNs. First, the conventional CNN models are analysed and evaluated for multisensor satellite image based land cover classification. It is observed from this analysis that since earth observation satellite images have usually larger spectral depths than normal camera images, conventional CNN models are not sufficient to exploit the high spectral information present in these images. However, exploiting the spectral information is crucial. In fact, similar to spatial domain, convolutions in the spectral domain are intuitive with high spectral dimension data. Convolutions in the spectral domain might also be an alternative to the process of feature selection. Based on this idea, chapter 5 presents two novel multisensor satellite image classification approaches that utilize CNNs in both spatial and spectral domains. The frameworks employ localized spectral convolutions along with localized spatial convolutions on permuted and stacked high

xiv

dimensional multisensor images. The study establishes the significance of performing localized spectral convolutions on permuted input in comparison to non-localized or non-permuted localized spectral convolutions. Two models, namely, a permuted local spectral CNN (Perm-LS-CNN) and a permuted local spectral-spatial CNN (Perm-LSS-CNN) are obtained.

For assessment of the performance of models (conventional and novel), PALSAR-2 PolSAR and Landsat-8 multispectral multisensor images over the Roorkee and Haridwar (Uttarakhand, India) study areas are considered. Since, PolSAR and multispectral images provide complementary information and they are widely used together in numerous applications. Five land cover classes are considered namely built-up, tall vegetation such as forests, bare soil, short vegetation such as agriculture crops, and water. The models are trained and tested on ground truth class data points measured directly on the terrain. Classification accuracy and generalization performance of key models are evaluated. Generalization performance is evaluated via visual inspection of classified images on selected well known regions in the study areas. In an attempt to quantize the generalization performance of classifiers, another parameter, AD is computed. AD is the difference between precision and recall, averaged over all the land cover classes. Comparison with conventional CNN models and other popular machine learning classifiers shows that the Perm-LSS-CNN model although has classification accuracies in the order of 90% (similar to other classifiers), but provides better generalization performance. Both indicators i.e. visual inspection and AD support to the better generalization performance of the Perm-LSS-CNN.

Chapter 6 presents a study in which an attempt is made; to analyse the conventional CNN and RNN models, and to develop novel CNN and RNN (in standalone as well as synergistic modes) models for agriculture crop classification with multisensor/multimodal satellite time series (MSTS) imagery. For this purpose, two sub tasks are set in this study. In the first sub task i.e. *sub task 1*, conventional CNN and RNN models are analysed and evaluated for the task of multisensor satellite time series (MSTS) images based crop classification. 1D, 2D, and 3D CNNs, RNNs, and combinations of these are considered for analysis. Then, based on the critical analysis of the working and performance of these conventional models, few novel models are proposed and developed in *sub task 2*.

The novel models involve permuted spectral and/or spatial convolutions and/or bi-directional recurrent layers. The Bidirectional RNNs capture the temporal profile and the permuted spectral and/or spatial CNNs capture the spectral and/or spatial profile of a particular crop. Bidirectional strategy in RNNs enables the influence of the past, present, and future inputs on the current output. Bidirectional RNNs are beneficial in time series images based crop

classification because crop classification can utilize the bidirectional process i.e. classification result can be observed from any time node. Also, the permuted spectral band stacking strategy developed in chapter 5 is also used in this study and presumed to improve the significance of the spectral information present in multisensor satellite time series images. In particular, a synergistic utilization of permuted spectral-spatial CNNs and Bidirectional RNNs helps in better exploitation of the multisensor satellite time series images.

Overall, 12 models, 6 conventional models namely; 1D-CNN-v1, 3D-CNN-v1, RNN-v1, 1D-CRNN-v1, 2D-CRNN-v1, and 3D-CRNN-v1, and 6 novel models, namely; Perm-1D-CNN-v1, Perm-3D-CNN-v1, Perm-RNN-v1, Perm-1D-CRNN-v1, Perm-2D-CRNN-v1 (also termed as Perm-LS-Bi-LSTM), and Perm-3D-CRNN-v1 (also termed as Perm-LSS-Bi-LSTM) are developed. For analysis and performance evaluation of the conventional and novel models, multimodal satellite (Sentinel-1 dual-PolSAR and Sentinel-2 multispectral) time series images over the Roorkee study area is considered. Wheat and sugarcane are the major crops grown in the area during the study period. Multisensor time series satellite images over a period of one crop (wheat) cycle (based on heuristic knowledge) are used. Also, to evaluate the robustness of the models under mixed land cover scenarios, land cover classes other than crop types such as forests and built-up areas are also included as output classes. Addition of these classes in the classification process resembles the actual scenario on the ground. Training and testing samples are once again measured and collected directly from the ground. Conventional and novel DNN based crop classification models are evaluated on the basis of classification accuracy and generalization performance. In terms of classification accuracy, Perm-1D-CNN-v1 achieves an accuracy of 99.70% which is the best among the models considered for comparison. In contrast, the generalization performance evaluated via visual inspection, indicated to a superior performance of the proposed Perm-3D-CRNN-v1 and Perm-1D-CRNN-v1 models as these models are able to capture the spectral, spatial, and temporal identity of land cover features more efficiently which results to improved generalization performance of these models.

Finally, *chapter 7* concludes the different studies reported in this thesis. The conclusions drawn from each study, contributions made by the studies and, future scopes of the studies are also listed in this chapter. In summary, based on the studies presented in chapter 3, 4, 5, and 6, it can be concluded that the novel CNN and RNNs models developed in these studies are high performance and scalable and are potentially automatable. The models have shown good generalization performance which makes the models robust and less prone to overfitting and also these models perform well in mixed class scenarios. Further studies are however required to establish these models in the satellite image processing community.

would not be possible or even instigated without the firm support of my lovely wife Ms. Deepmala Phartiyal who had been my friend and invisible guide, always caring and nurturing when I was stressed or tensed. I admire her for unconditional love and patience during all these years.

Writing these few lines give me utmost pleasure as these lines are dedicated to my friends; Akhil, Nikhil, Vishal, Deepmala, Prateek, Lokesh, Pankaj, Ganesh, Tarun, Avantika, Navin, Eeshan, Rakesh, Alaknanda, Shilpi, Hitesh, Sachin, Ashutosh, Bharat, Ayush, Mandakini, Purnima, shambhaivi, Ruchi, Ayushi, Avantika. Without these guys, I would not have been able to withstand the pressure of this journey. Akhil, Nikhil, Vishal, Avantika, Deepmala were the sparkles who kept me cheered up all along. I would like to thank everyone whom I met and was part of journey during this period.

I would like to thank Prabhu, Bhagwaan, Ishwar, and God for keeping a blessing eye on me. This journey would not have been smooth if not for his showers of wisdom.

It would be impossible to acknowledge all the people who contributed either technically or by offering their friendship towards the completion of this work. I am thankful to all those who were associated with my research journey directly or indirectly and to whom I forgot to mention here unknowingly. My thesis has been a wonderful experience and I cherish the educational and personal growth it has provided to me for my future endeavors.

Gopal Singh Phartiyal

xxii

TABLE OF CONTENTS

xxx

xxxii

AL	Area limit
ALOS	Advanced land observing satellite
AMSR-E	Advanced microwave scanning radiometer for earth observation
ANNs	Artificial neural networks
BBOR	Bounding box overlap ratio
Bi-LSTM	Bidirectional long short term memory
BOA	Bottom-of-Atmosphere
BS	Bare Soil
CERES	Clouds and the earth's radiant energy system
CNN	Convolutional neural network
Corr_co	Co-polarized correlation
Corr_cross	Cross-polarized correlation
DBNs	Deep belief networks
DCAE	Deep conventional auto encoder
DEM	Digital elevation models
DESIS	DLR earth sensing imaging spectrometer
Di	Dihedral
DL	Deep learning
DNN	Deep neural network
DT	Decision tree
DV	Detected village
EM	Electromagnetic
EnMAP	Environmental mapping and analysis program
FC	Fully Connected
FIR	Far infrared
FP	Flat plate
FP	False positive
FWHM	Full width and half maximum
GA	Genetic algorithm
GFD	Generalized Freeman Durden decomposition

GOES	Geostationary operational environmental satellite
GPS	Global positioning system
GRU	Gradient recurrent unit
GTI	Ground truth image
HD	Horizontal dipole
HH	Horizontal transmit and horizontal receive
HSI	Hyperspectral imagery
HV	Horizontal transmit and vertical receive
IAEG-SDG	Inter-agency and expert group on SDG indicator
IFOV	Instantaneous field of view
InSAR	Interferometric synthetic aperture radar
IOA	Individual overall accuracy
ISRO	Indian space research organization
JAXA	Japan aerospace exploration agency
LandSat	Land remote-sensing satellite
LC	Land Cover
LIDAR	Light detection and ranging
LSTM	Long short term memory
LWIR	Long wave infrared
M	Mean
MAP	Maximum a posterior
MLP	Multilayer perceptron
MODIS	Moderate-resolution imaging spectroradiometer
MOPITT	Measurements of Pollution in the Troposphere
MRI	Magnetic resonance imaging
MS	Multispectral
MSI	Multispectral instrument
MSS	Multispectral scanner
MTS	Multisensor time series
MWIR	Mid wave infrared
m-Y4D	Modified Yamaguchi four component decomposition
NAIP	National agricultural innovation project
NASA	National aeronautics and space administration

NDVI	Normalized difference vegetation index
NED	Normalized Euclidean distance
NIR	Near infrared
NISAR	NASA ISRO SAR
NSCM	Normalized signature correlation mapper
OA	Overall accuracy
OLI	Operational land imager
PALSAR	Phased array type L-band synthetic aperture radar
PCA	Principal component analysis
Perm-LS-CNN	Permuted local spectral convolutional neural network
Perm-LSS-CNN	Permuted local spectral-spatial convolutional neural network
PolSAR	Polarimetric synthetic aperture radar
ppi	Pixels per inch
PS	Polarization signature
PSCF	Polarization signature correlation features
RBF	Radial basis function
RBMs	Restricted Boltzmann machine
RCS	Radar cross section
ReLU	Rectified linear unit
RF	Random forest
RNN	Recurrent neural network
ROI	Regions of interest
RS	Remote sensing
S	Standard deviation
SAE	Stacked auto encoder
SAR	Synthetic aperture radar
SCM	Signature correlation mapper
SI	Separability index
SLC	Single look complex
SNR	Signal to noise ratio
SONAR	Sound navigation ranging
SRTM	Shuttle radar topography mission
SV	Short vegetation

SVM	Support vector machine
SWIR	Short wave infrared
TIRS	Thermal infrared sensor
TN	True negative
TP	True positive
TRMM	Tropical rainfall measuring mission
TV	Tall vegetation
UN	United Nations
VD	Vertical dipole
VH	Vertical transmit and horizontal receive
VV	Vertical transmit and vertical receive

Chapter 1

INTRODUCTION

The planet is continually being observed and imaged by satellites. Before 1972, satellites weren't designed to study or monitor earth's surface. Instead, they were mainly used for military missions. Satellite based earth imaging was commercialized in 1984, but faced many funding issues. This led to the passing of the land remote sensing policy act of 1992. The new law embraced the value of satellite based earth imaging technology, promoted research and public sector applications, and allowed commercial companies to launch satellites, acquire images of earth's surface, and utilize it or sell it. As a result, earth observation satellites specifically intended for non-military applications were designed for long-term global observations of the land surface, oceans, biosphere, and atmosphere.

In the last decade, there has been a surge in the availability of high quality (spatial, spectral, and temporal), open-source/economical earth observation satellite images. Today, more than 600 satellites provide earth observation data, mostly in the form of images. Satellite images are available at different spatial, spectral, radiometric, and temporal resolutions. Also, satellite images are available in different modalities such as surface reflectance (multispectral, superspectral, or hyperspectral), backscatter (active microwave), emission (passive microwave, thermal), polarization (linear, circular, or elliptical), and others. These satellite images are used for numerous civil and defence applications. Popular categories of satellite image applications include land cover monitoring, agriculture crop monitoring, disaster monitoring, environment health and earth surface dynamics monitoring, and mineral exploration. These applications vary from being performance-critical such as land cover classification to being time-critical such as disaster management. Therefore, different applications require focus on different characteristics (spatial, spectral or temporal) of satellite images. The different characteristics, modalities, and applications of satellite images are further discussed in section 1.1.

Processing satellite images for utilization in different applications is challenging. Literature reports hundreds of approaches or techniques available for satellite image processing for various applications. The popular satellite image processing techniques are decision tree, random forest, maximum likelihood, support vector machines, artificial neural networks, k-nearest neighbours [1], logistic regression [2], distance measure [3], K-means, ISODATA, and more. These techniques are briefly discussed in section 1.2. However, it is observed that development of techniques, that can explore different properties (different resolutions or different

modalities) offered by different satellite images, is tedious and crafty let alone, keeping track of more resourceful satellite images every day and improvising the techniques accordingly is not an easy task. Therefore, it is expected that new techniques should possess universal applicability, scalability, automatability, and generalization. Generalization is the measure of the ability of machine learning techniques on how well these techniques perform on 'out of sample' data. The challenges with currently popular satellite image processing techniques and new expectations are discussed in section 1.3.

Deep learning (DL) with deep neural networks (DNNs) have shown substantial potential in satellite image processing applications [4]. The deep learning theory is a universally applicable approach which can map multiple levels of information to represent complex relationships between data [5]. Various aspects of deep learning favourable to satellite image processing application are discussed in section 1.4.

However, the deep neural networks are still young, promising, and growing techniques in the satellite image processing community and further exploration of the potential of these models is motivating. The present chapter attempts to establish the need of a comprehensive study on the potential, significance, advantage and drawbacks with using deep neural network models for extraction of information from satellite images. This chapter also highlights some major research gaps in the development of more efficient and generalized deep neural network models. Based on the research gaps, new objectives are resolved and are discussed in this chapter.

1.1. Background: Satellite Images

Satellite imagery depicts the earth's surface at various spectral, temporal, radiometric, and spatial resolutions, as determined by the sensing device (sensor) on board each collection system, and the orbital path of its reconnaissance platform [6]. Satellite sensors acquire information reflected or emitted from an object in the form of electromagnetic radiation and then the information about the state and condition of an object is derived by interpreting this received electromagnetic signal. The collected information (in the form of images) indicates about the physical, chemical, and biological properties of observed object which may be soil [7], [8], ice, snow [9], [10], water [11], wind [12], vegetation [13] or any other. The goal of collecting the information is to understand how electromagnetic energy interacts with the earth's surface objects. The information can further be used for crucial military and civil applications such as monitoring; vegetation coverage [14], soil moisture [15], disasters [16], sea-ice fluctuation, sea level and sea surface temperatures [17], [18], ships in sea [19], concentrations of atmospheric

gases, and many other parameters. These observations are applied to meteorological forecasting, environmental monitoring, disaster monitoring, and cartographic purposes.

Satellite images can be gathered using different ways of "sensing" the object: (i) by reflection, (ii) by emission, and (iii) by combined emission-reflection.

1.1.1. Satellite image: Characteristics

A satellite image has four different characteristics namely spatial resolution, spectral resolution, radiometric resolution, and temporal resolution. The most common of these is spatial resolution, but other aspects are also very important for image interpretation.

A. Spatial Resolution

Spatial resolution is the fineness of detail in the image and is measured in pixel per inch (*ppi*). The more pixels per inch, better the resolution. Generally, an image with a higher resolution produces a better printed image quality. Spatial resolution helps to identify the minimum separation at which objects appear independent and isolated. It is measured in millimetres on the photograph or in meters on the ground, and it depends on the focal length of the camera and its height above the ground. The most common definition of sensor spatial resolution is the size of the projected instantaneous field of view (IFOV) on the ground (d), computed as in equation (1.1).

$$d = 2h \tan(\frac{\text{IFOV}}{2}) \tag{1.1}$$

Where d is the distance on the ground for each information unit (pixel) and h the height of the observing sensor. IFOV is defined as the angular section observed by the sensor, in radians, at a given moment in time.

The spatial resolution of optical-electronic sensors depends on orbit height, speed of data collection, and the number of detectors. As for the antenna sensors such as SAR, their resolution depends on the aperture angle, height of the platform, and the wavelength at which they work. The larger the aperture angle, the lower the height and the wavelength, and the more detailed the resolution. Figure 1.1 shows a subset of image acquired via the Landsat-8 sensor which has 30 m spatial resolution. It means that a single pixel represents an area of 900 square meters (30m ×30m) on the ground. The spatial resolution of imaging sensors improved significantly over time as shown in Figure 1.2 and is usually inversely related to acquisition period and coverage area.
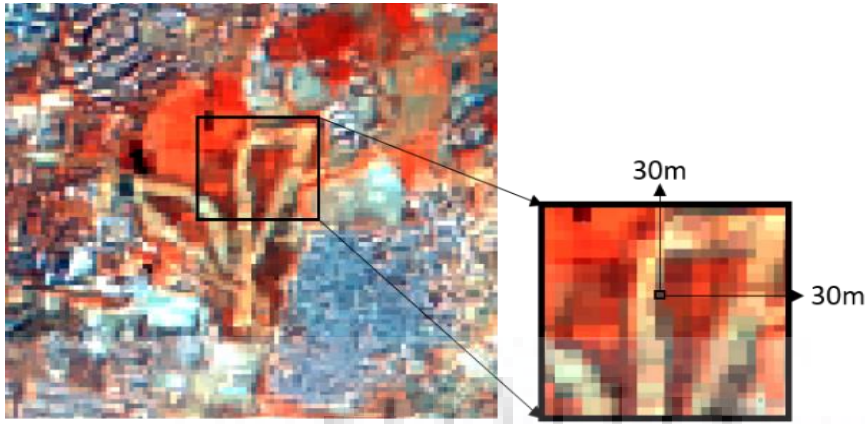
Figure 1.1 Landsat-8 pixel of 30m spatial resolution.
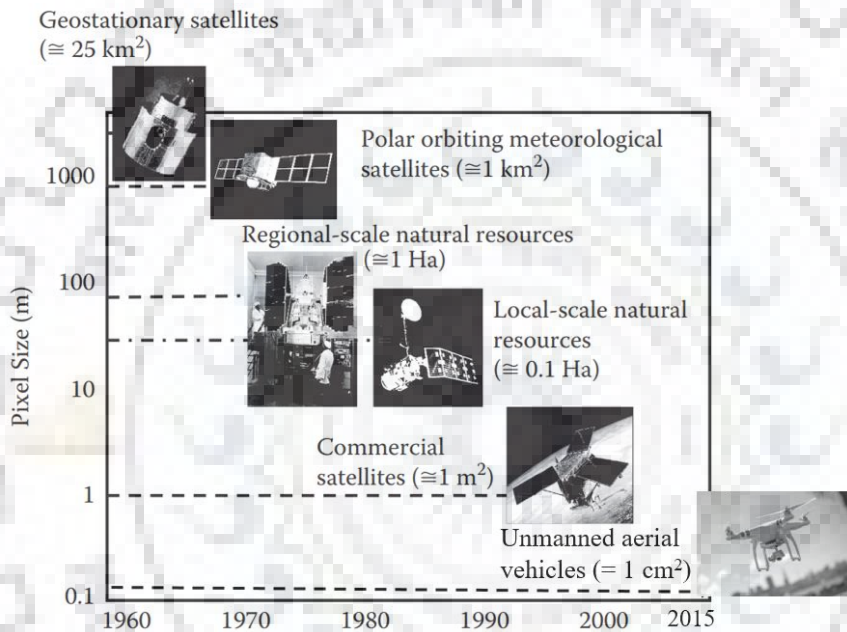
Figure 1.2 Improvements in image spatial resolutions over time.

B. Spectral Resolution

Spectral resolution refers to the number of bands provided by the sensor, as well as their spectral bandwidths. Generally speaking, a sensor will provide better discrimination capacity as more bands are acquired. Multispectral satellite imaging systems record energy in multiple bands of the electromagnetic spectrum. Table 1.1. briefly summarizes the categorization of the electromagnetic (EM) spectrum in terms of spectral bands and spectral bandwidths. The nominal size of a band may be large (i.e. coarse), as with the Landsat MSS near-infrared band 4 (800 – 1,100 nm) or relatively smaller (i.e. fine), as with the Landsat MSS band 3 (700 – 800 nm). Thus, Landsat MSS band 4 detectors recorded a relatively large range of reflected near-infrared radiant flux (300 nm wide) while the MSS band 3 detectors recorded a much reduced range of near-infrared radiant flux (100 nm wide). A hyperspectral satellite imaging sensor typically acquires images in hundreds of spectral bands. For example, the earth observing-1 Hyperion has 220

bands in the region from 356 to 2,577 nm spaced just 10 nm apart based on the full width and half maximum (FWHM) criteria.

C. Radiometric Resolution

The radiometric resolution of an imaging system (including satellite sensors) is described as its ability to discriminate very slight differences in energy. The finer the radiometric resolution of a sensor, the more sensitive it is to detecting small differences in reflected, scattered, or emitted energy. Figure 1.3 demonstrates the concept of radiometric resolution. In Figure 1.3, the image represented with 1 bit has only 2 levels of intensities whereas the image represented with 16 bits has 65,536 levels of intensities. While the arrangement of pixels describes the spatial distribution of a structures present in an image, the radiometric characteristics represent information about the interaction of the structure with electromagnetic energy. Every time an image is acquired on film or sensor, its sensitivity to the change in magnitude of the electromagnetic energy determines its radiometric resolution.

D. Temporal Resolution

Temporal resolution is defined as the amount of time needed to revisit and acquire data for the exact same location. When applied to satellite imaging based earth observation, temporal resolution is controlled by a sensor's swath width alongside the orbital parameters of its sensing platform; together these parameters determine the periodicity (return interval) at which a fixed area on the ground is successively imaged by a space borne sensor. It is also important to bear in mind that the effective temporal resolution depends on the atmospheric conditions because optical and thermal sensors cannot detect surfaces under cloud cover. That is why in some areas the availability of useful images can be notably lower that the acquisition cycle of the sensor (tropical wet regions, for instance). To somehow alleviate this problem, modern sensors have the capacity to make observations pointing the sensor at different angles, facilitating the collection of data over a zone of interest from an adjacent orbit.

Temporal resolution is a primary consideration when selecting an image product for researchers studying time series behaviour. Sensors with higher temporal resolution or more extensive imaging archives are however, most likely to afford cloud-free and seasonally favourable images, making them attractive options for all investigators. Figure 1.4 depicts the temporal revisit period of an arbitrary satellite as a typical example.

Table 1.1 Broad categorization of the electromagnetic spectrum in terms of spectral bands and their spectral bandwidths [21].

EM Radiation Regions	Wavelength	Applications
Gamma-ray	$<10^{-3}$ nm	Food purification, medical hygienic and nuclear medicine
X-ray	0.1 to 10^{-3} nm	Medical diagnostics and industry area
Ultra-violet	3 to 400 nm	For computer/lab imaging, fluorescence microscopy and food (meat) quality evaluation
Visible	Red: 610-700 nm Orange: 590-610 nm Yellow: 570-590 nm Green: 500-570 nm Blue: 450-500 nm	Environment monitoring, land surface dynamics assessment, mineral exploration
Near infrared	Near infrared (NIR): 0.7 to 1.5 μm Short wave infrared (SWIR): 1.5 to 3 μm Mid wave infrared (MWIR): 3 to 8 μm Long wave infrared (LWIR): 8 to 15 μm Far infrared (FIR): >15 μm	Earth observation Vegetation monitoring
Microwave	P band: 30-100 cm L band: 15-30 cm S band: 7.5-15 cm C band: 3.8-7.5 cm X band: 2.4-3.8 cm Ku band: 1.7-2.4 cm K band: 1.1-1.7 cm	Imaging for cloud free and weather independent acquisition. Location tracking
Radio	Very short wave, VHF: 1-10 m Short wave, HF: 10-100 m Medium wave, MF: 0.1-1 km Long wave, LF: 1-10 km Very long wave, VLF: 10-100 km	Communication operations, Astronomical applications, magnetic resonance imaging (MRI)

2-bit 4 levels

4-bit 16 levels

8-bit 256 levels

16-bit 65,536 levels
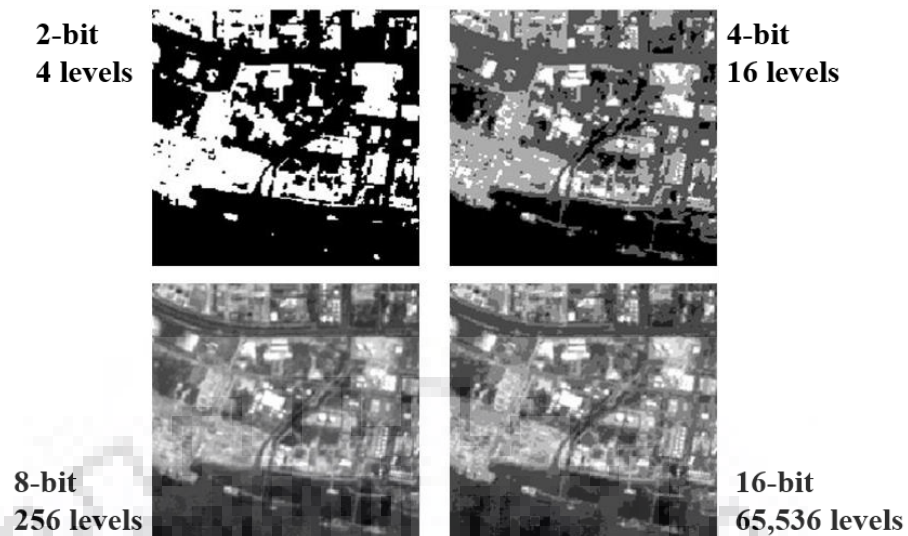
Figure 1.3 Demonstration of the concept of radiometric resolution. The larger the bit depth, the higher the radiometric resolution (more number of intensity levels). Source: https://faq.spatial.farm/cn/technology/

November, 2017

December, 2017

January, 2017

30 days

T_1

T_2

T_3
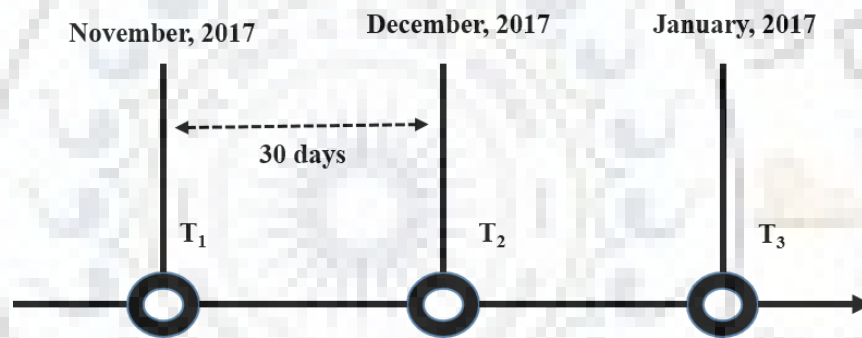
Figure 1.4 The temporal resolution of a satellite imaging system refers to how often it records imagery of a particular area. This example depicts the systematic collection of data every 30 days, presumably at approximately the same time of day.

1.1.2. Satellite images: Acquisition types and sensors

One of the most common ways to classify sensor systems is on the basis of the mechanism they use to detect electromagnetic energy. In this regard, two broad groups of sensors: i) Passive, which simply collect radiation derived from external sources, and ii) Active, in which the sensor system is able to provide both, the energy source and detect reflection of the energy from the observed surfaces. Obviously, active sensors are more flexible because they are independent of the sensor–earth illumination conditions. A further distinction among passive sensors is to divide them according to the method used to record the incoming energy.

Passive Imaging: Passive remote imaging works in one of the following three ways: i) By observing how radiation changes as it passes through objects (such as sunlight passing through a thin clouds), ii) By observing radiation emitted by an object of interest (such as the emission of

7

infrared radiation by earth), or iii) By observing radiation reflected or scattered by an object of interest (such as a visible satellite image taken from space). These three methods take advantage of naturally occurring radiation (such as sunlight or infrared radiation emitted by objects) and as such, provide a wealth of information all around us. For example, human eyes serve as types of passive remote imaging devices - our eyes capture scattered and emitted visible light, which the human brain process into images. In contrast, satellite imaging devices such as the Landsat series satellites or the Sentinel-2 satellite constellation have aboard them sensors than can record electromagnetic radiation beyond visible light and provide a great deal of information about our environment.

Active Imaging: Active remote imaging devices work by, first emitting a signal, and then recording the "return" of the emitted signal from an object of interest. It includes, observing how long it took the signal to return, the magnitude of the returned signal, and changes in the signal properties such as frequency shifts or orientation (phase, polarization) of the signal. In nature, bats make extensive use of their natural active remote imaging system. They emit ultrasonic 'squeaks' which bounce off from objects around them including the 'moths'. The bats use this active imaging method to locate moths for food. A bat can 'see' (locate) a moth flying in the dark by processing to the changes in the returned ultrasonic signal from the moths. Whales and dolphins use a similar system to augment their other senses for communications and hunting. Their system is termed as sound navigation and ranging or more popularly as 'SONAR'. Satellite-based active remote imaging systems include RADARS (an acronym for radio detection and ranging) and LIDARS (an acronym for light detection and ranging) which work by emitting electromagnetic radiation (either radio waves, in the case of radar, or a laser beam for the LIDAR system) in the direction of the object to be sampled, and then utilizing a parabolic dish to collect the return (called 'backscatter') radiation for processing by computers.

Popular satellite imaging modalities

A. *Optical imaging for earth observation*

Imaging in the optical region of the electromagnetic (EM) spectrum implies use of visible, near infrared and short-wave infrared sensors to form images of the earth's surface by detecting the solar radiation reflected or scattered from objects on the earth's surface. Different earth surface objects reflect and absorb differently at different wavelengths. Thus, these objects can be identified, differentiated, or detected based on their spectral reflectance signatures recorded by satellite optical sensors. Optical imaging systems are majorly classified into four types namely

panchromatic, multispectral, superspectral, and hyperspectral depending on the number of spectral bands used in the imaging process.

In panchromatic imaging systems, the imaging sensor is a single wide spectral band (channel) detector sensitive to radiation covering the visible light. A panchromatic image (grey scale image) is one band spread over a couple of hundred nanometres bandwidth. The bandwidth enables the panchromatic sensor to hold a high signal-noise, making the panchromatic image available at a higher spatial resolution. The physical quantity being measured is the apparent brightness of the targets. The spectral information or "colour" of the targets is lost. Recent examples of panchromatic imaging systems are: WorldView-PAN, IKONOS-PAN, GeoEye-PAN, SPOT high resolution visible (HRV-PAN) etc.

The image acquisition sensor in a multispectral imaging system is a multiband/multichannel detector. Each channel is sensitive to radiation within a narrow wavelength band. The resulting image is a multilayer image which contains both the brightness and spectral (colour) information of the objects being observed. Examples of multispectral systems are: Landsat MSS/TM/ETM/ETM+/OLI, Sentinel-2 A/B multispectral instrument (MSI), etc. The band specifications of few multispectral imaging sensors are discussed here.

For example, Landsat-8 operational land imager (OLI) and thermal infrared sensor (TIRS) images consist of nine spectral bands with a spatial resolution of 30 meters for Bands 1 to 7 and 9. Band 1 (ultra-blue) is useful for coastal and aerosol studies. Band 9 is useful for Cirrus cloud detection. The resolution for Band 8 (panchromatic) is 15 meters. Thermal bands 10 and 11 are useful in providing more accurate surface temperatures and are collected at 100 meters. Approximate scene size is 170 km north-south by 183 km east-west. Landsat-8 band specifications are listed in Table 1.2. The sensors on Landsat-9 (launch ready mid-2021) are being designed as improved copies of Landsat-8. Sentinel-2 MSI covering 13 spectral bands (443–2190 nm), with a swath width of 290 km and a spatial resolution of 10 m (four visible and near-infrared bands), 20 m (six red edge and shortwave infrared bands) and 60 m (three atmospheric correction bands). Sentinel-2 band specifications are listed in Table 1.3. The Sentinel-2A MSI data has spectral bands very similar to Landsat 8 (excluding the thermal bands of Landsat 8's thermal infrared sensor).

Table 1.2 Landsat-8 (multispectral) OLI and TIRS band specifications [22].

Bands	Bandwidth (µm)	Resolution (m)	Potential Applications
Band 1- Coastal aerosol	0.43-0.45	30	Coastal and aerosol studies
Band 2- Blue	0.45-0.51	30	Bathymetric mapping, distinguishing soil from vegetation and deciduous from coniferous vegetation
Band 3- Green	0.53-0.59	30	Emphasizes peak vegetation, which is useful for assessing plant vigour
Band 4- Red	0.64-0.67	30	Discriminates vegetation slopes
Band 5- Near Infrared (NIR)	0.85-0.88	30	Emphasizes biomass content and shorelines
Band 6- SWIR 1	1.57-1.65	30	Discriminates moisture content of soil and vegetation, penetrates thin clouds
Band 7- SWIR 2	2.11-2.29	30	Improved moisture content of soil and vegetation and thin cloud penetration
Band 8- Panchromatic	0.50-0.68	15	Sharper image definition
Band 9- Cirrus	1.36-1.38	30	Improved detection of cirrus cloud contamination
Band 10- Thermal Infrared (TIRS) 1	10.6-11.19	100	Thermal mapping and estimated soil moisture
Band 11- Thermal Infrared (TIRS) 2	11.50-12.51	100	Improved thermal mapping and estimated soil moisture

Table 1.3 Sentinel-2 band specifications [23].

Bands	Wavelength (µm)	Resolution (m)	Potential Applications
Band 1- Coastal aerosol	0.443	60	Aerosol scattering
Band 2- Blue	0.490	10	Bathymetric mapping
Band 3- Green	0.560	10	Bathymetric mapping
Band 4- Red	0.665	10	Bathymetric mapping
Band 5- Vegetation Red Edge	0.705	20	Vegetation classification
Band 6- Vegetation Red Edge	0.740	20	Vegetation classification
Band 7- Vegetation Red Edge	0.783	20	Vegetation classification
Band 8- NIR	0.842	10	chlorophyll, biomass
Band 8A- Vegetation Red Edge	0.865	20	Vegetation classification
Band 9- Water Vapour	0.945	60	Detection of thin cirrus
Band 10- SWIT Cirrus	1.375	60	Water vapour absorption
Band 11- SWIR	1.610	20	lignin, starch, and forest above biomass
Band 12- SWIR	2.190	20	Distention to live biomass

A superspectral imaging sensor has typically 15 to 50 spectral channels. The bands have narrower bandwidths, enabling the finer spectral characteristics of the targets to be captured by the sensor. An example of superspectral imaging system is: moderate-resolution imaging spectroradiometer (MODIS) which have 36 spectral bands and its band specifications are listed in Table 1.4.

Table 1.4 MODIS sensor band specifications and primary usage [25].

Band	Bandwidth[1]	Spectral Radiance[2]	Required SNR[3]	Potential Applications
1	620-670	21.8	128	Land/Cloud/Aerosols Boundaries
2	841-876	24.7	201	
3	459-479	35.3	243	Land/Cloud/Aerosols Properties
4	545-565	29.0	228	
5	1230-1250	5.4	74	
6	1628-1652	7.3	275	
7	2105-2155	1.0	110	
8	405-420	44.9	880	Ocean Colour/ Phytoplankton/
9	438-448	41.9	838	Biogeochemistry
10	483-493	32.1	802	
11	526-536	27.9	754	
12	546-556	21.0	750	
13	662-672	9.5	910	
14	673-683	8.7	1087	
15	743-753	10.2	586	
16	862-877	6.2	516	
17	890-920	10.0	167	Atmospheric Water Vapour
18	931-941	3.6	57	
19	915-965	15.0	250	
20	3.660-3.840	0.45 (300 K)	0.05	Surface/ Cloud Temperature
21	3.929-3.989	2.83 (335 K)	2.00	
22	3.929-3.989	0.67 (300 K)	0.07	
23	4.020-4.080	0.79 (300 K)	0.07	
24	4.433-4.498	0.17 (250 K)	0.25	Atmospheric Temperature
25	4.482-4.549	0.59 (275 K)	0.25	
26	1.360-1.390	6.00	150	Cirrus Cloud Water vapour
27	6.535-6.895	1.16 (240 K)	0.25	
28	7.175-7.475	2.18 (250 K)	0.25	
29	8.400-8.700	9.58 (300 K)	0.05	Cloud Properties
30	9.580-9.880	3.69 (250 K)	0.25	Ozone
31	10.780-11.280	9.55 (300 K)	0.05	Surface/ Cloud Temperature
32	11.770-12.270	8.94 (300 K)	0.05	
33	13.185-13.485	4.52 (260 K)	0.25	Cloud Top Altitude
34	13.485-13.785	3.76 (250 K)	0.25	
35	13.785-14.085	3.11 (240 K)	0.25	
36	14.085-14.385	2.08 (220 K)	0.35	

[1]Bands 1 to 19 are in nm; Bands 20 to 36 are in μm

[2]Spectral Radiance values are (W/m2 – μm-sr)

[3]SNR = Signal- to – Noise Ratio

[4]NE (Delta)T = Noise-equivalent temperature difference

Note: Performance goal is 30-40 % better than required

Table 1.5 Hyperion sensor specifications [26].

Specification	Setting
Spatial resolution	30 m
Swath width	7.75 km
Spectral channels	220 unique channels:
	VNIR (70 channels, 356 nm-1058 nm)
	SWIR (172 channels, 852 nm-2577 nm)
Spectral bandwidth	10nm (nominal)
Digitization	12 bits
Signal –to – noise ratio (SNR)	161 (550 nm)
	147(700 nm)
	110 (1125 nm)
	40 (2125 nm)

A hyperspectral imaging system, also known as an "imaging spectrometer" collects information in about a hundred or more contiguous spectral bands. The goal is to collect precise spectral information about objects of interest such as minerals, oil, gas, forests or others. This information enables better characterisation and identification of objects. Hyperspectral images have potential applications in fields as coastal management (e.g. monitoring of phytoplankton, pollution, bathymetry changes) [24], forestry, water spectroscopy, mineral exploration, precision agriculture (e.g. monitoring the types, health, moisture status and maturity of crops). An example of a hyperspectral system is: Hyperion which has 220 spectral bands. Hyperion sensor specifications are listed in Table 1.5.

B. *Microwave imaging for earth observation*

Electromagnetic radiation in the microwave region is also used in imaging to provide useful information about the earth's surface and objects on it. There are various types of microwave imaging systems developed to capture information in different forms.

A microwave radiometer is a passive sensing device which records the microwave emissions emitting naturally from the earth's surface. The sensing technology can be used in many applications such as measuring and monitoring sea surface temperature or land surface temperature etc.

A radar altimeter sends out pulses of microwave energy and records the energy scattered back from the object on the earth surface. The height of the object from a certain point on earth's surface can be measured based on the time taken by the energy signal to return.

A wind scatterometer can be used to measure wind speed and direction over the ocean surface. It sends out pulses of microwaves along several directions and records the magnitude of

the signals backscattered from the ocean surface. The magnitude of the backscattered signal is related to the ocean surface roughness, which in turns is dependent on the sea surface wind condition, and hence the wind speed and direction can be derived from platforms to generate high resolution images of the earth surface using microwave energy.

Synthetic aperture radar

In synthetic aperture radar (SAR) imaging, microwave pulses are transmitted by an antenna towards the earth surface at an oblique angle and the microwave energy scattered back towards the antenna (backscatter) is measured and used to analyse earth surface features. The SAR imaging system form an image by utilising the magnitude and time delay of the backscattered signals. Mathematically, this backscattered signal is represented using radar cross section (RCS) σ, which is defined as the ratio between the incident and received signal intensity. The microwave energy beam sent out by the antenna towards the earth's surface illuminates an area on the ground. The illuminated area is termed as the antenna's "footprint". The recorded or received signal strength depends on the microwave energy backscattered from the objects inside this footprint. In real aperture radar imaging, the ground resolution is limited by the size of the microwave beam sent out from the antenna. Finer details on the ground can be resolved by using a narrower beam. The beam width is inversely proportional to the size of the antenna i.e. the shorter the antenna, the wider is the beam or footprint and the longer the antenna, the narrower the beam or footprint. Increasing the length of the antenna will decrease the width of the footprint. It is not feasible for a spacecraft to carry a very long antenna which is required for high resolution imaging of the earth surface. To overcome this limitation, a SAR capitalises on the motion of the space craft to emulate a large antenna from a small antenna it actually carries on board.

SAR polarimetry

In general, when an EM wave is emitted from a source (e.g. a radar antenna), it propagates in all directions (a spherical wavefront), with a specific field strength and a phase in each direction. At a long distance from the antenna, it can be assumed that the wavefront of the emitted EM wave lies on a plane, rather than on the surface of a sphere. Polarization is an important property of an EM wave. Polarization refers to the alignment and regularity of the electric and magnetic field components of the EM wave, in a plane perpendicular to the direction of propagation. In particular, EM wave can be characterized by the behaviour of the electric field vector as a function of time. The propagation of an EM wave is illustrated in Figure 1.5. A scatterer (object on the surface of earth in this case) can change the polarization of the EM wave incident upon it. The polarization of the scattered EM wave can be different from the polarization of the incident

13

EM wave. Usually, a radar antenna is often designed to receive a predefined set of polarization components of the scattered EM wave simultaneously. For example, the horizontal (H) and vertical (V) parts of an antenna can receive the two orthogonal components of the incoming EM wave, and the system electronics keep these two signals separate.
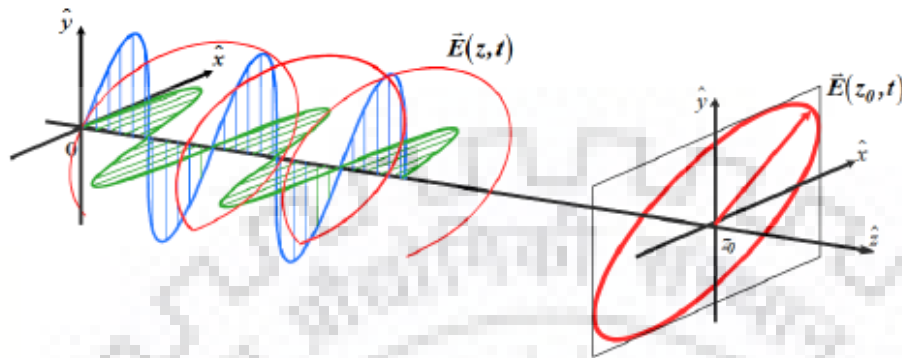
Figure 1.5 Illustration of the propagation of an electromagnetic plane wave. The electric field vector has horizontal (green) and vertical (blue) components, which combine to yield the net electric field vector (red) [27].

Denoting the transmit and receive polarizations by a pair of symbols, a radar system using H and V linear polarizations can thus have the following channels:

- HH - for horizontal transmit and horizontal receive,
- VV - for vertical transmit and vertical receive,
- HV - for horizontal transmit and vertical receive, and
- VH - for vertical transmit and horizontal receive.

The first two of these polarization combinations are referred to as like-polarized or co-polarized, because the transmit and receive polarizations are the same. The last two combinations are referred to as cross-polarized because the transmit and receive polarizations are orthogonal to one another [28].

A radar system can have different levels of polarization complexity. Popular polarization configurations are listed here (though not limited to).

- Single polarized (single channel) - HH or VV or HV or VH
- Dual polarized (dual channel) - HH and HV, VV and VH, or HH and VV
- Quador full polarized (four channels) - HH, VV, HV, and VH

A quadrature polarized (i.e. fully polarimetric) radar uses these four polarizations, and measures the phase difference between the channels as well as the magnitudes. Some dual polarized radars also measure the phase difference between channels, as this phase plays an important role in polarimetric information extraction.

When a horizontally polarized wave is incident upon a target, the backscattered wave can have contributions in both horizontal and vertical polarizations. The same applies to a vertically

14

polarized incident wave. As the horizontal and vertical components form a complete basis set to describe an electromagnetic wave, the backscattering properties of the target can be described by a scattering matrix, S as.

$$[S] = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \tag{1.2}$$

The four elements of the scattering matrix are complex, and can be obtained from the magnitudes and phases measured by the four channels of the receiving antenna of a polarimetric radar.

The different polarization channels and derivatives represents different scattering mechanisms, and hence offer different information of objects as mentioned in Table 1.6. For example, HH-VV polarization gives double-bounce scattering, which is usually obtained from buildings in a landscape consisting of a variety of land covers.

Historical, current and future SAR sensors:

Spaceborne SAR sensors have been in use for more than 40 years. The first SAR sensor was launched on June 28, 1978, on board the NASA's Seasat satellite. The SAR sensor was developed and deployed, aimed at monitoring oceanographic phenomena. Seasat carried an HH-polarized L-band SAR sensor mounted at a fixed angle to observe the global 'surface wave fields' and polar sea ice conditions as part of its sensor suite. Even though the Seasat's SAR sensor operated for only 106 days, the mission was deemed an extensive success, demonstrating a SAR's capability to monitor both ocean and land surface.

Since then, SAR imaging has made a long and successful journey. Several SAR sensors with ever-improving imaging capabilities have been used by organizations, communities, nations collectively ensuring continuous coverage of the earth and availability of SAR data [29]. Unfortunately, this constellation of SAR systems comes with a downside. The SAR sensors launched by the various agencies vary widely in their sensor configurations. The data from the different sensors are not always directly comparable with each other.

Table 1.6 Polarimetric parameters.

Parameter	Application (Discrimination)
HH-VV	Double bounce scattering
HV	Volume scattering
HH+VV	Surface scattering

SAR sensors transmit energy in one of the microwave frequency bands shown in Table 1.7. Roughly speaking, radar systems use frequencies from 1 to 90 GHz. This frequency range is

subdivided into multiple frequency bands, as shown in the first column of Table 1.7. These bands were defined initially according to the different equipment(s) needed to generate and detect these particular frequencies. However, now they are understood as the equivalent of colours in the visual range. Unfortunately, the latter-based naming scheme (Ka-band to P-band) was used during World war II and never deemed necessary to be modified since then. This nomenclature sometimes leads to confusion among new users of SAR.

Not all of the microwave bands shown in Table 1.7 used are for SAR imaging. While some experimental airborne Ka- and Ku- band SAR systems exist, civilian spaceborne sensors have been using the frequency bands ranging from P- to X-band.

The operating wavelength of a SAR sensor is linked intrinsically to its transmitted signal's penetration capabilities. It means that longer wavelength signals (e.g., signals at L- and P-band) penetrate deeper into land covers and earth surface compared to shorter wavelength signals i.e. at C- or X-band. Hence, the applications supported by a SAR sensor depends on the frequency band used.

Table 1.7 also summarizes typical applications of SAR as a function of the frequency band. It shows that sensors at X-band predominantly are used for built-up and infrastructure monitoring. Due to the X-band's higher resolution capabilities, SAR sensors at this frequency find broad applications in surveillance and tracking. These SAR sensors are also often used in the monitoring of industrial installations. Due to its limited penetration into vegetation covers, X-band is rarely used to characterise forest canopies or monitor activity underneath vegetation.

With the predominant number of SAR sensors operating at this frequency range, C-band sensors have been the workhorse of SAR based monitoring over the last 30 years. C-band data can be seen as a good compromise between X-band and the longer wavelength (e.g. L-band) sensor classes because of its moderate- to high-resolution capabilities and increased vegetation penetration. In comparison to X- band SARs in terms of swath widths, C- band SARs typically allow for wider swath imaging, lending its applicability to regional and global scale. While C-band has improved canopy penetration capabilities, its signals will typically not penetrate vegetation completely. C-band is of limited use for monitoring activity underneath vegetation canopy layers, especially in regions with dense vegetation.

While S-band SAR sensors were used rarely in earth observation in the past, this frequency will have increased usage soon. NovaSAR-S, a S-band SAR sensor, was launched in September 2018. NovaSAR-S will provide moderate-resolution SAR data. This data will help explore S-band data's performance for applications such as disaster monitoring, crop monitoring, forest monitoring (temperate and rainforests), and land cover mapping. More interesting to most users will be the upcoming NASA-ISRO SAR satellite, NISAR. In addition to an L-band SAR,

16

NISAR will carry a fully polarimetric S-band SAR. While NISAR's S-band coverage will likely not be global, all data will be freely and openly available to the SAR science and applications community.

While most historical SAR systems operated in C-band, future SAR sensors are focused on the L-band frequency range. Although L-band SARs do not provide the high-resolution capabilities of shorter wavelength SARs, their ability to penetrate vegetation holds several advantages for earth observation. With a higher likelihood of seeing the ground underneath vegetation cover, L-band SARs are useful for mapping activity underneath canopies such as flooding or troop movements. Due to the high penetration into vegetation covers, L-band SAR can also characterise canopy structure, especially in dense vegetation. The higher canopy penetration ability of L-band SARs is also advantageous to interferometric SAR (InSAR) users in terms of achieving higher interferometric coherence and better deformation tracking.

P-band SAR sensors are currently under development. The major issue with P-band SAR's data is 'ionospheric distortions'. These ionospheric distortions can adversely hamper the quality of data. However, recent developments in ionospheric correction strategies have allowed spaceborne P-band SAR missions to go forward. The first spaceborne P-band SAR—ESA's BIOMASS mission is planned to launch in 2021 and will focus on mapping the status and dynamics of earth's forests.

A list of the most relevant past, current, and future SAR platforms is provided in Table 1.8. The sensors are sorted by their period of functioning. For each instrument, sensor's operating wavelength, supported polarization modes, resolution and size of image products, repeat cycle, and data access means are listed.

Table 1.7 Designation of microwave bands [30].

Band	Frequency (GHz)	Wavelength (cm)	Typical Application
Ka	27-40	1.1-0.8	Rarely used for SAR (airport surveillance)
K	18-27	1.7-1.1	Rarely used (H_2O absorption)
Ku	12-18	2.4-1.7	Rarely used for SAR (satellite altimetry)
X	8-12	3.8-2.4	High-resolution SAR (urban monitoring; ice and snow, little penetration into vegetation cover; fast coherence decay in vegetated areas)
C	4-8	7.5-3.8	SAR workhorse (global mapping; change detection; monitoring of areas with low to moderate vegetation; improved penetration; higher coherence); Ice, ocean, maritime navigation
S	2-4	15-7.5	Little but increasing use for SAR-based earth observation; agriculture monitoring (NISAR will carry an S-band channel; expands C-band applications to higher vegetation density)
L	1-2	30-15	Medium resolution SAR (Geophysical monitoring; biomass and vegetation mapping; high penetration; InSAR)
P	0.3-1	100-30	Biomass. First P-band spaceborne SAR will be launched ~2020; vegetation mapping and assessment. Experimental SAR.

Table 1.8 Brief list of few past, current and upcoming spaceborne SAR sensors with their properties [30].

Sensor	Lifetime	Wavelength /Frequency	Polarization	Resolution	Frame Size	Repeat Cycle	Access
Seasat	1978	L-band λ =24.6 cm	HH	Az:25m Rg:25m	100km	-	Free& Open
ERS-1	1991-2001	C-band λ =05.6 cm	VV	Az:6-30m Rg:26m	100km	35days	Restrained
JERS-1	1995-1998	L-band λ =24.6 cm	HH	Az:18m Rg:18m	75km	44days	Restrained
ERS-2	1995-2011	C-band λ =05.6 cm	VV	Az:6-30m Rg:26m	100km	35days	Restrained
ENVISAT	2002-2012	C-band λ =05.6 cm	HH, VV, VV/VH HH/HV,VV/VH	Az:28m Rg:28m	100km	35days	Restrained

18

ALOS-1	2006-2011	L-band λ =24.6 cm	FBS: HH, VV FBD: HH/HV, HH/VH PLR: HH/HV/VH/VV ScanSAR: HH, HV	FBS:10*10m FBD:10*10m PLR:30*10m ScanSAR:100m	FBS:70km FBD:70km PLR:30km ScanSAR:250-350km	46days	Free & Open
Radarsat-1	1995-2013	C-band λ =05.6 cm	HH	Standard: 25x28m Fine: 9x9m Wide1: 35x28m Wide2: 35x28m ScanSAR: 50x50-100x100m	Standard: 100km Fine: 45km Wide1: 165km Wide2:150km ScanSAR: 305-510km	24 days	1995-2008: Restrained 2008-2013: Commercial
TerraSAR-X TanDEM-X	2007-2010	X-band λ = 03.5cm	Single: HH, VV Dual: HH/VV, HH/HV, VV/VH Twin: HH/VV, HH/VH, VV/VH	Spotlight: 0.2x1.0-1.7x3.5m Stripmap: 3x3m ScanSAR: 18-40m	Spotlight: 3-10km Stripmap: 50x30km ScanSAR: 150x100-200x200km	11 days	Application-dependent; restrained scientific, commercial
Radarsat-2	2007-	C-band λ = 05.6cm	Single: HH, VV, HV, VH Dual: HH/HV, VV/VH Quad: HH/HV/VH/VV	Spotlight: ~1.5m Stripmap: ~3x3-25x25m ScanSAR: 35x35-100x100m	Spotlight: 18x8km Stripmap: 20-170m ScanSAR: 300x300-500x500km	24 days	Commercial
COSMO-SkyMed	2007-	X-band λ = 03.5cm	Single: HH, VV, HV, VH Dual: HH/HV, HH/VV, VV/VH	Spotlight: ≤1m Stripmap: 3-15m ScanSAR: 30-100m	Spotlight: 10x10km Stripmap: 40x40km ScanSAR: 100x100 -200x200km	Satellite: 16 days Constellation: ~hrs	Commercial; limited proposal-based scientific
XSAR RISAT-2	2009-	X-band λ = 03.1cm	Single: HH, VV, HV, VH Dual: HH/HV, HH/VV, VV/VH Quad: HH/HV/VH/VV	Spotlight: ≤1m Stripmap: 3m Mosaic: 1.8m ScanSAR: 8m	10x10-120x120km	14 days	Restrained scientific

ALOS-2 PALSAR-2	2014-	L-band λ = 24.6cm	Single: HH, VV, HV, VH Dual: HH/HV, VV/VH Quad: HH/HV/VH/VV	Spotlight: 1x3m Stripmap: 3-10m ScanSAR: 25-100m	Spotlight: 25x25km Stripmap: 55x70-70x70km ScanSAR: 355x355km	14 days	Commercial; limited proposal-based scientific
Sentinel-1	2014-	C-band λ = 05.6cm	Single: HH, VV Dual: HH/HV, VV/VH	Stripmap: 5x5m Interferometric Wide Swath (IW): 5x20m Extra Wide Swath (EW): 20-40m	Stripmap: 375km IW: 250km EW: 400km	Satellite: 12 days Constellation: 6 days	Free & Open
SAOCOM	2018-	L-band λ = 24.6cm	Single: HH, VV Dual: HH/HV, VV/VH Quad: HH/HV/VH/VV	Stripmap: 10x10m Top SAR: 100x100m	Stripmap: >65km Top SAR: 320km	Satellite: 16 days Constellation: 8 days	TBD
NovaSAR-S S-SAR	2018-	S-band λ = 11.75cm	Single: HH, VV Dual: HH/HV, VV/VH Quad: HH/HV/VH/VV	ScanSAR = 20m Maritime Surveillance = 30m Stripmap = 6m ScanSAR Wide = 30m	Stripmap = 15-20 km ScanSAR Wide = 150lm Maritime Surveillance = 750km		TBD
PAZ SAR	2018-	X-band λ = 03.5cm	*See TerraSAR/Tan DEM-x	*See TerraSAR/Tan DEM-x	*See TerraSAR/Tan DEM-x	11 days	Commercial
RCM	2019	C-band λ = 05.6cm	Single: HH, VV, VH, HV Dual: HH/HV, VV/VH, HH/VV Compact Quad	Very high, high, medium, and low-res modes (3-100m)	20x20-500x500km	Satellite: 12 days Constellation: ~hrs	TBD
NISAR	2021	L-band λ = 24.6cm S-band	Single: HH, VV, VH, HV	3-20m (mode dependent)	250km	12 days	Free & Open

		$\lambda = 11.75$cm	Dual: HH/HV, VV/VH, HH/VV Quad				
BIOMASS	2021	P-band $\lambda = 70.0$cm	Quad	≤60x50m	160km	17 days	Free & Open
TanDEM-L	2023	L-band $\lambda = 24.6$cm	Single, dual, quad modes	12*12 m	350 km	Satellite: 16 days Constellation: 8 days	Free & Open
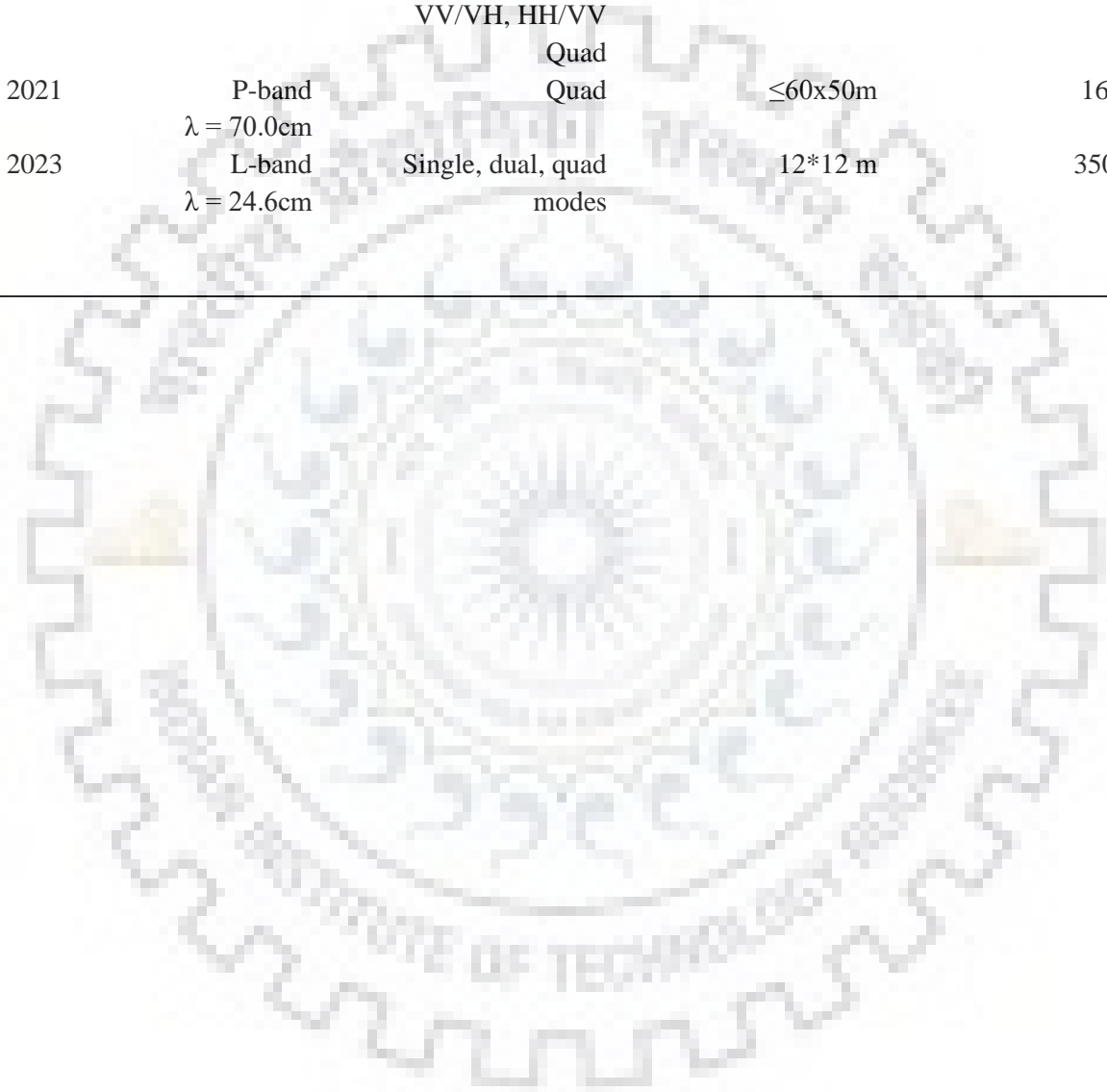
1.2. Satellite Images: Summary and Context to the Thesis

It is evident from section 1.1 that satellite images are critical to earth monitoring and are the most techno-economical means of doing so. Following are key takeaways from satellite images based earth monitoring.

Satellite images provide wide coverage, timely response, reliable and consistent information, and high quality [31] earth observation data. Satellite image are available at different characteristics i.e. spatial, spectral, and temporal resolutions. Also, satellite images are available in various modalities such as surface reflectance (multispectral, or hyperspectral), backscatter (active or passive microwave), brightness-temperature (microwave, thermal), polarization (linear, or circular), and others. These different satellite image characteristics and modalities are useful in standalone or synergistic fashion for various applications. This thesis focuses on efficiently utilizing these characteristics and modalities, in standalone or synergistic fashion, for different applications.

On top of this, satellite images are available in large amounts. Specifically, in the last decade, there has been a surge in the availability of high quality (spatial, spectral, and temporal), open-source or economical earth observation satellite images. Today, more than 600 satellites provide earth observation data, mostly in the form of images. Although there are various satellite image modalities available today, but PolSAR (backscatter) and multispectral (surface reflectance) are the most techno-economical, widely used satellite image modalities. This thesis focuses on studies based on these two image modalities for different applications.

Popular satellite image applications include land cover classification and mapping, agriculture crop monitoring, disaster monitoring, environment health and earth surface dynamics monitoring, and mineral exploration. These applications vary from being performance-critical such as land cover classification to being time-critical such as disaster management. Therefore, different applications require focus on one or more characteristics (spatial, spectral or temporal) or modality (spectral reflectance, backscatter, brightness-temperature) of satellite images. Although, *single sensor or unimodal* satellite images are able to provide satisfactory quality spatial, spectral, or temporal information but a trade-off is always at play. For example, if a satellite image has very good spatial resolution, then the satellite sensor is ought to take more time to gather this detailed spatial information than the sensor which provides satellite image with lower spatial information. In short, higher the spatial resolution, lower the temporal resolution, keeping all other aspects same. Similar trade-off holds for spectral and spatial

characteristics of satellite images. These trade-offs hinder the ability of a single sensor satellite image to be sufficient for every application.

Alternatively, use of multimodal and multisensor satellite images address the issues associated with the use of single sensor images. Use of multisensor images improves the spatial, spectral, and temporal characteristics simultaneously and therefore improving the application performance. Moreover, use of multisensor satellite images also address to some other issues associated with unimodal images. For example, multispectral images (modality: surface reflectance) are less noisy and easy to interpret but suffer from issues like obstruction due to cloud cover or no night-time imaging. On the other hand, PolSAR images (modality: backscatter) are weather-independent (can be captured at any time) but suffer from issues such as noise (speckle, layover, foreshortening etc.) and image interpretation. Use of multisensor images resolves these issues to some extent. Therefore, use of ***multisensor and/or multimodal*** satellite images for many applications is not uncommon. Also, with the availability of multiple sources of information for any particular application, it is important to exploit all the available sources efficiently and effectively. Therefore, this thesis also focuses to exploit multisensor satellite images for land cover classification, agriculture crop classification, and disaster monitoring.

Satellite time series images are set of images of the same area captured at different time stamps using the same satellite imaging sensor. Multisensor time series images include time series images from more than one satellite sensor. Both, single sensor and multisensor satellite time series images have proved potential in various applications such as change detection, agriculture crop monitoring, etc. This thesis focuses to exploit multisensor satellite time series images for agriculture crop classification.

In summary, unimodal or single sensor satellite images, multimodal or multisensor satellite images, and satellite time series images have advantages, disadvantages, and varied applicability for earth observation. This thesis focuses to exploit all the three categories for different applications.

1.3. Satellite Image Processing: Recent Challenges and Expectations

Processing satellite images for utilization in different application is challenging. In general, images acquired and utilized for earth observation are different from normal images because; - *first,* earth observation images contain significantly higher spatial information. A single image may contain spatial information about from a few kilometres to few hundred kilometres of the earth's surface, and –*second*, earth observation images also contain relatively higher spectral information. It means unlike the normal camera images which have three spectral bands i.e. red,

green, and blue, spectral bands in these images may vary from tens to hundreds. These unique characteristics of earth observation images make them challenging to handle and require special approaches to process. Satellite based earth observation images also possess these properties and are equally challenging to be utilized in desired applications such as land cover classification or agriculture crop monitoring. The challenges can be divided into three categories for the sake of clarity i.e. challenges with single sensor satellite image processing, challenges with multisensor satellite image processing, and challenges with satellite time series image processing. These challenges are briefly discussed here.

Single sensor and unimodal satellite images have been used in numerous earth observation applications. Among a wide variety of applications, land use land cover or land cover (LC) classification is one of the most important and mundane task. Traditional techniques have been successful in performing land cover classification with various unimodal satellite images [32]. Few of the popular image classification techniques are decision trees, random forest [33], maximum likelihood, support vector machines, artificial neural networks, k-nearest neighbours [1], logistic regression [2], distance measure [3], K-means, ISODATA, and more. These techniques have been widely used in satellite image based land cover classification for many years and have performed well [32]. However, these techniques possess certain limitations such as requirement of a feature extraction process, high-dimensional data inoperability, sub-optimal classification, performance saturation with increase in availability of trainable data [5], [34] or due to local minima, and low generalization performance [35]. Generalization is the ability of a classifier or classification technique to perform well outside the training sample space. Researchers in the satellite image processing community are always looking forward to techniques that can process the high dimensional satellite images and possess abilities like scalability, automatability, generalizability, and high-performance in mixed land cover scenarios. For example, satellite polarimetric synthetic aperture radar (PolSAR) earth observation data is weather independent data but it suffers from issues such as overlapping of class observables (backscatter) due to weak sensor isolation or speckle noise and therefore often results in sub-optimal land cover classification with traditional classifiers in mixed class scenarios. Mixed class scenarios refer to study regions where land cover classes considered for classification are small, unorganized, and are scattered randomly. Due to this landscape two or more different land covers appear similar in a satellite image. For example, built-up areas having building structures close to each other and are randomly oriented (with respect to SAR sensor flight path), might show volume scattering along with double bounce scattering thus behaving as mixed class. Numerous techniques are available to perform land cover classification with satellite PolSAR images.

Decomposition based [13], [36]–[41], factorization based [42], and machine learning based techniques [43]–[46] are most popular with satellite PolSAR image based land cover classification. Decomposition and factorization based methods mostly rely on exploring physical scattering properties [47]–[49] and texture information [50], [51] in SAR images. Also, these techniques are mainly based on specific problems and characteristics of SAR data sources. On the other hand, machine learning techniques are more generic in nature but they suffer from limitations such as overfitting and low generalization. Similar issues are observed with other modalities of satellite images. Therefore, more efficient (high performance), scalable and robust techniques are to be explored in order to achieve improved classification performances. Similar to land cover classification, other satellite image based applications such as crop and disaster monitoring can benefit from more advanced techniques. Moreover, traditional approaches with unimodal optical images also suffer from similar challenges.

Alternatively, synergistic exploitation of multimodal or multisensor satellite images is also a challenge for traditional techniques. Numerous multisensor satellite image processing techniques have been developed and have proved successful to some extent [52]–[56]. However, the popular multisensor satellite image processing techniques such as data fusion techniques are data and application intensive and tend to loose information during the fusion process [52], [57]. Machine learning techniques such as neural networks are also popular multisensor satellite image processing techniques but they suffer from overfitting and low generalization performance issues [58]. Therefore, there are still possibilities for improvements in the current multisensor satellite image processing techniques and there is scope for new ones.

Furthermore, efficient processing of satellite time series images is also critical for many applications [59]–[64]. For example, crop type identification and classification is a major step during agriculture crop monitoring [65], [66]. Accurate crop classification requires synergistic exploitation of spatial, spectral and temporal profiles of crops [67]. Numerous techniques have been reported so far, for satellite time series image based crop classification such as in [68], but certain limitations still exist such as presence of other land covers (built-up, forests, or grasslands) in and around the croplands under study. Therefore, it is still a challenge for existing techniques to synergistically exploit the spatial, spectral, and temporal information of a crop present in unimodal/single-sensor or multimodal/multisensor satellite time series images and provide accurate and reliable crop classification results [60].

In summary, literature reports plethora of techniques available for satellite image processing be it for unimodal or single sensor images, multimodal or multisensor satellite images, or satellite time series images. Overall, it is observed that development of techniques that can efficiently exploit different characteristics and/or different modalities offered by satellite images

is tedious and crafty. On top of it, keeping track of more resourceful satellite images every day and improvising the techniques accordingly is not an easy task at all. Based on these observations, it is safe to say that any novel technique, approach, method, or model should prefer to possess qualities such as high performance, universal applicability framework, scalability, automatability, and generalization. Presence of these attributes will help any technique, approach, method, or model to keep up with the advancements in satellite images.

1.4. Motivation: Deep Learning in Satellite Image Processing

In this direction of improvements in satellite image processing techniques, deep learning with deep neural networks have shown substantial potential [4]. In the recent years, many satellite image based applications have used deep neural networks and their popularity is still growing [69], [70]. The deep learning theory is a universally applicable approach which can map multiple levels of information to represent complex relationships between data [5]. The deep neural network models are high performance, scalable, automatable, and can learn directly from raw data (less feature handling) [35]. These are the abilities researchers were looking towards from new techniques as discussed in section 1.3. For example, Figure 1.6 indicates the power of deep learning to learn directly from data. The deep learning based models improve their performance with increase in availability of training data in contrast to other machine learning techniques that are unable to learn beyond a certain point no matter how much training data is available. This means that, the more the training data, the better the deep learning based model performs. Another advantage of deep learning is its independency from hand-crafted features. Traditional machine learning techniques require hand-crafted features (not strictly) for learning. Figure 1.7 illustrates this advantage. These attributes or abilities of deep learning based models placed them in prominent position to be widely used for satellite image processing.

Although numerous techniques and methods are available that are based on deep learning and DNNs, there is a general framework at play. Figure 1.8 illustrates the general framework of DNN based models for satellite image analysis and processing. The framework includes three main components; an input data, a core deep networks, and expected output. In practice, the input–output pairs are dependent on a particular application. When the input–output pairs are properly defined, the intrinsic and natural relationship between the input and output is then constructed by a deep network composed of multiple levels of linear and nonlinear operations and a deep learning algorithm. When the core deep network has been well trained, it can be employed to predict the expected output of an input (test) whose output is never seen by the core deep network.

26
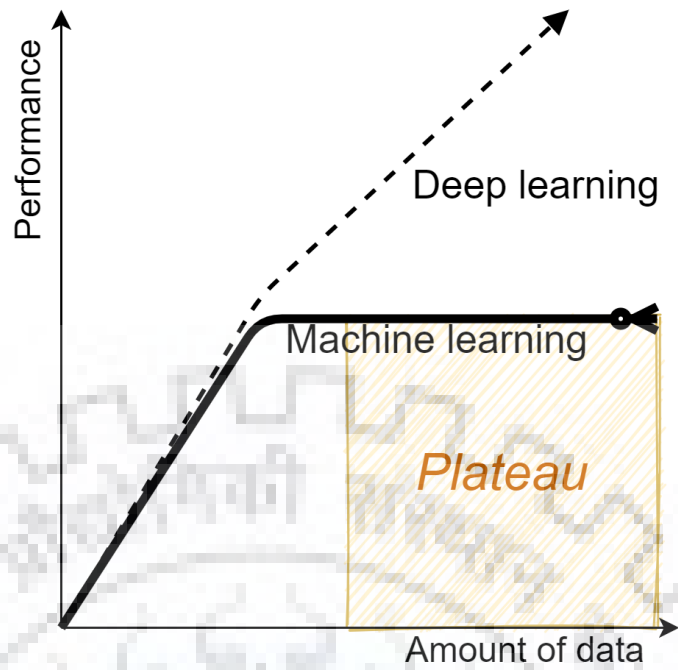
Figure 1.6 Comparison of performances of traditional machine learning and deep learning techniques with respect to amount of training data availability.
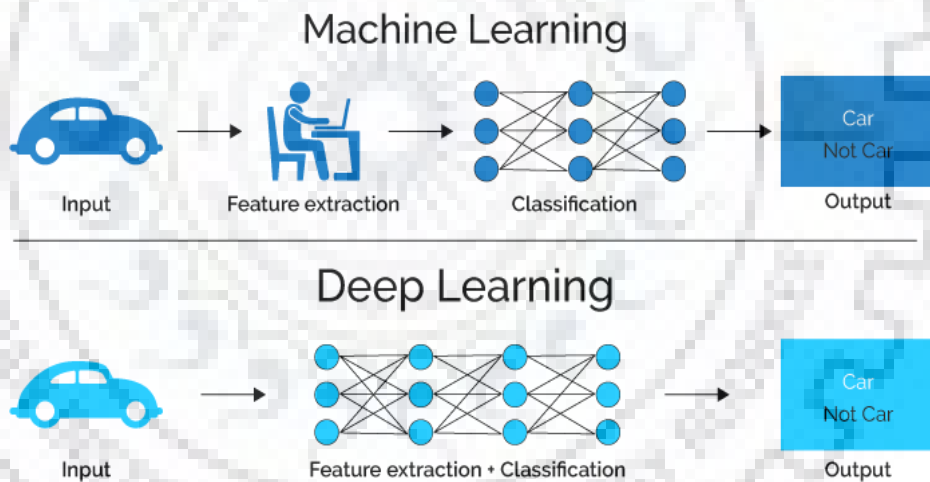
Figure 1.7 Figure illustrates that deep learning models can directly learn from raw data. The feature selection/extraction process is therefore redundant and sub-optimal in these models [71].
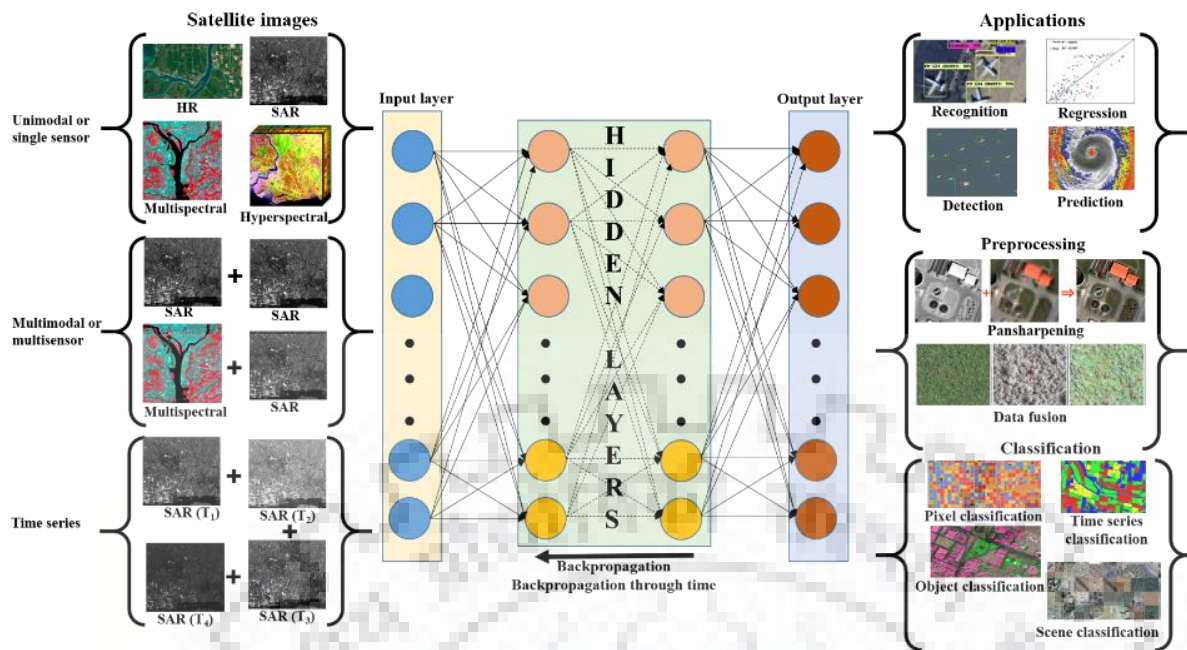
Figure 1.8 A general framework in DNN based models for satellite image analysis and processing.

For example, Convolutional neural networks or CNNs are recently the most widely used deep network models in satellite image based land cover classification [5]. CNNs have the ability to extract complex features directly from data [35]. Another advantage of focusing on developing CNNs for satellite image processing is that CNNs can directly utilize multi-sensor high dimensional satellite images with efficient performance [72]. Unlike the conventional data fusion methods, this reduces the need of fusing the multisensor images together before using it further. Furthermore, CNNs and recurrent neural networks or RNNs are the most recent and trending deep neural network frameworks for satellite time series image based agriculture crop monitoring applications. Since crop classification is the most challenging as well as important task in many crop monitoring applications, both CNN and RNN models have been developed in, standalone and synergistic modes for agriculture crop classification by many researchers [73], [74]. The aim of these models is to capture the spatial attributes of a particular crop type with CNNs and/or its temporal attributes with RNNs. These studies reported good classification results and promoted further studies. A detailed review on the recent trends and applications of CNNs and RNNs based satellite image processing is provided in chapter 2 here. The review involves major satellite image applications developed on uni- and multi-modal, single- and multi- sensor, single time stamp and time series satellite images with the help of deep learning and DNNs. The review indicates to the significant success and advantage of deep learning methods, especially CNNs and RNNs in the field of satellite image processing. However, the CNN and RNNs deep neural network models are still young, promising, and growing techniques in the satellite image

processing community and further exploration of the potential of these models, especially CNNs is motivating. Studies such as, improving the generalization performance of CNN based land cover classification models by focusing on the spectral attributes of satellite images or, exploring synergistic use of bi-directional RNNs with spectrally focussed CNNs for agriculture crop monitoring applications may be beneficial.

1.5. Research Gaps in Deep Learning based Satellite Image Processing

A brief literature review is carried out in the field of deep learning based satellite image processing. Chapter 2 provides brief review, description and discussions on various novel deep learning based satellite image processing techniques developed for various applications. Although, these techniques have proved their potential and have gained popularity over the years, there are areas that require special attention. There are areas of research in deep learning and DNNs based satellite image processing where improvements are worth mentioning. Few of these are:

- Although numerous DNN models have been developed so far for satellite image analysis and processing, we believe there is more to explore.
- Improving the performance of current DNN models or development of novel DNN models in mixed land cover scenarios is still a challenge.
- **Spectral information in satellite images is under exploited with current DNN techniques:** Earth observation satellite images contain relatively higher spectral information in comparison to normal camera images. This means unlike the normal camera images which have three spectral bands i.e., red, green, and blue, spectral bands in these images may vary from tens to hundreds. These characteristics of earth observation images makes them challenging to handle and require special attention. Current DNN techniques, especially CNNs and RNNs mostly focus on spatial and temporal characteristics of satellite images during satellite image processing for different applications. Focussing spectral information along with spatial/temporal information might be beneficial.
- **Need more attention on development of more generalized and robust DNN based satellite image processing techniques:** It is intuitive to say that similar to conventional neural networks, DNNs also suffer from overfitting and low generalizability. Special focus is required to improve the generalization ability of DNNs in mixed land cover scenarios. Ensemble training is key to achieve good generalization in DNN models. However, literature reveals lack in ensemble training based DNN models for satellite image processing. Also, visual inspection based evaluation of generalization ability of DNN models needs to be promoted.
- Multisensor satellite images, especially optical and PolSAR complement each other in various land cover applications. There is a lack of CNN and RNNs based models which can exploit multisensor, especially PolSAR and optical images

directly at various temporal, spectral and spatial resolutions for land cover applications. Special focus is required on the complementary spectral information present in high dimensional multisensor satellite images.

- Although, processing time series satellite imagery with CNNs and RNNs for crucial applications such as agriculture crop classification is gaining popularity, the weightage on the exploitation of temporal, spatial, and spectral information via CNNs and RNNs is not well investigated and needs attention. Also, the element of multisensor makes it more interesting and motivating.
- Since PolSAR and multispectral satellite images are the most abundant, economical, and popular modalities, special focus is required to develop CNN and RNN methodologies which can utilize these modalities for different land cover applications.

The above-mentioned research gaps are addressed (to some extent) in this thesis and various DNN models are proposed that focus on exploiting the information present in satellite images more carefully.

1.6. Thesis Objective

The main objective of this thesis is; the study and critical analysis of the performance of conventional CNN and RNN models and further development of the novel CNN and RNN models to exploit the spectral and/or spatial and/or temporal information present in unimodal/single senor or multimodal/multisensor satellite images. Special focus during the studies is to develop novel deep neural network (DNN) models (common term for CNN and RNN based models) that can achieve better generalization performance in mixed class scenarios.

In the attempt, four tasks are set to forward the stated objective. The tasks are based on; type(s) of satellite image(s) used, application for which the task is realized, multi-modality images involvement, and presence of time series information. The tasks and sub-tasks set to be carried out are:

Task 1: Development of CNNs and polarization signatures based novel approach for improved land cover classification with PolSAR satellite images.

Sub-tasks:

- Critical analysis of polarization signatures as potential feature for PolSAR image based land cover classification
- Development of novel class boundary estimation algorithm for optimal utilization of polarization signature based features to improve land cover classification accuracy.
- Development of novel CNNs based approach that directly utilizes polarization signatures to achieve land cover classification.

Task 2: Development of a novel approach using polarization signatures and CNNs for detection of human settlements affected during floods.

Sub-tasks:

- Development of a polarization signatures and CNN based model to extract built-up areas from PolSAR image.
- Development of a detection approach to detect villages from built-up areas using image morphology and *apriori* information.
- Mapping flood affected villages using image morphology.

Task 3: Development of novel spectral and spectral-spatial CNN models for multisensor (PolSAR and multispectral) images based land cover classification.

Sub-tasks:

- Study of conventional CNN models for multisensor (PolSAR and multispectral) image based land cover classification.
- Development of novel CNN models to achieve better generalization performance by efficiently utilizing spectral and spatial information from multisensor image.

Task 4: Development of novel CNN, RNN, and CNN-RNN models for multisensor satellite time series images based crop classification.

Sub-tasks:

- Study of conventional CNN and RNN models for multisensor satellite time-series images based crop classification.
- Development of novel CNN and RNN models with improved performance in multisensor satellite time-series images based crop classification.by synergistically exploiting spectral, spatial, and temporal profiles of crops.

1.7. Framework of the Research

Although each task follows a separate methodology, all the four tasks can be consolidated into a generic framework which is shown in Figure 1.9. The generic framework encompasses majorly five components namely: problem or objective identification, experiment setup, data preprocessing, model development and training, and model deployment, results, and discussion. The order might differ for different tasks based on whether the solution is empirical or analytical.

Problem or objective identification

In this part of the research, either a problem at hand is explored or an objective is laid out. In the former, the challenges with current applications are critically analysed first, then new technologies, materials, and methods are searched in order to find motivation to solve the problem. Task 1, 3, and 4 listed in section 1.6 are set out in this manner. Whereas in the latter,

first an objective or vision is set and then the goal is to attain the objective with best efficiency. Task 2 is set out in this manner. In both the cases, knowledge of the background and extensive literature review is a prerequisite. Various challenges and objectives identified during the present research are listed in section 1.4 and 1.5 respectively.

Experiment setup

Once either the problem is identified or an objective is set, the next step is to set up an experiment. The experiment setup includes materials such as acquiring datasets, selecting study areas, and collecting ground truth information and set of hypothetical methods which are suitable candidates for providing a solution are investigated. Since the present research is focussed on development of DNN based models for satellite image (especially PolSAR and multispectral) based applications, a set of satellite images (PALSAR-2, Landsat-8, Sentinel-1, and Sentinel-2) are considered for study. Study areas which are easily accessible and potential for subject of interest are identified. The study areas considered are mostly in the Haridwar district of Uttarakhand, India. These study areas considered have mixed land cover scenario and are interesting to evaluate any novel techniques. Extensive ground truth is collected in the study areas over the period of research. Hypothetical techniques considered in the present research are CNNs and RNNs.

Data preprocessing

The data considered in the present research are PolSAR and multispectral satellite image. Preprocessing PolSAR data involves calibration, speckle filtering, and terrain correction whereas preprocessing multispectral data involves atmospheric correction, cloud removal, and resampling. Both the datasets are preprocessed before to be used for various applications. Data preparation such as rescaling and stacking of PolSAR and multispectral images for multisensor data analysis is also carried out under this component. Further, the various CNNs and RNNs models take input data in particular formats, the pre-processed data is restructured accordingly under this component.

Model development and training

The general approach in this component of the research framework involves development of novel approach, algorithm, or model based on hypothetical methods and nature of data. In the present research, various CNN and RNN models are either investigated and newly developed for various applications. As per CNNs, one, two, and three dimensional convolution based novel models are proposed in task 1, 2, 3, and 4. Also, LSTM and GRU recurrent cell based RNN

32

models are developed as well for task 4. Further for task 4, synergistic exploitation of both the CNNs and RNNs is investigated and novel CNN-RNN models are proposed. The developed models are trained with ground truth information and under different scenarios.

Model deployment, results, and discussion

The final component of any research is the deployment of developed model and evaluation of results obtained. Accuracy and generalization are key evaluation parameters considered in the present research. Results are evaluated on multiple study areas. Generalization performance is evaluated via visual inspection of the results especially in task 3, and 4.
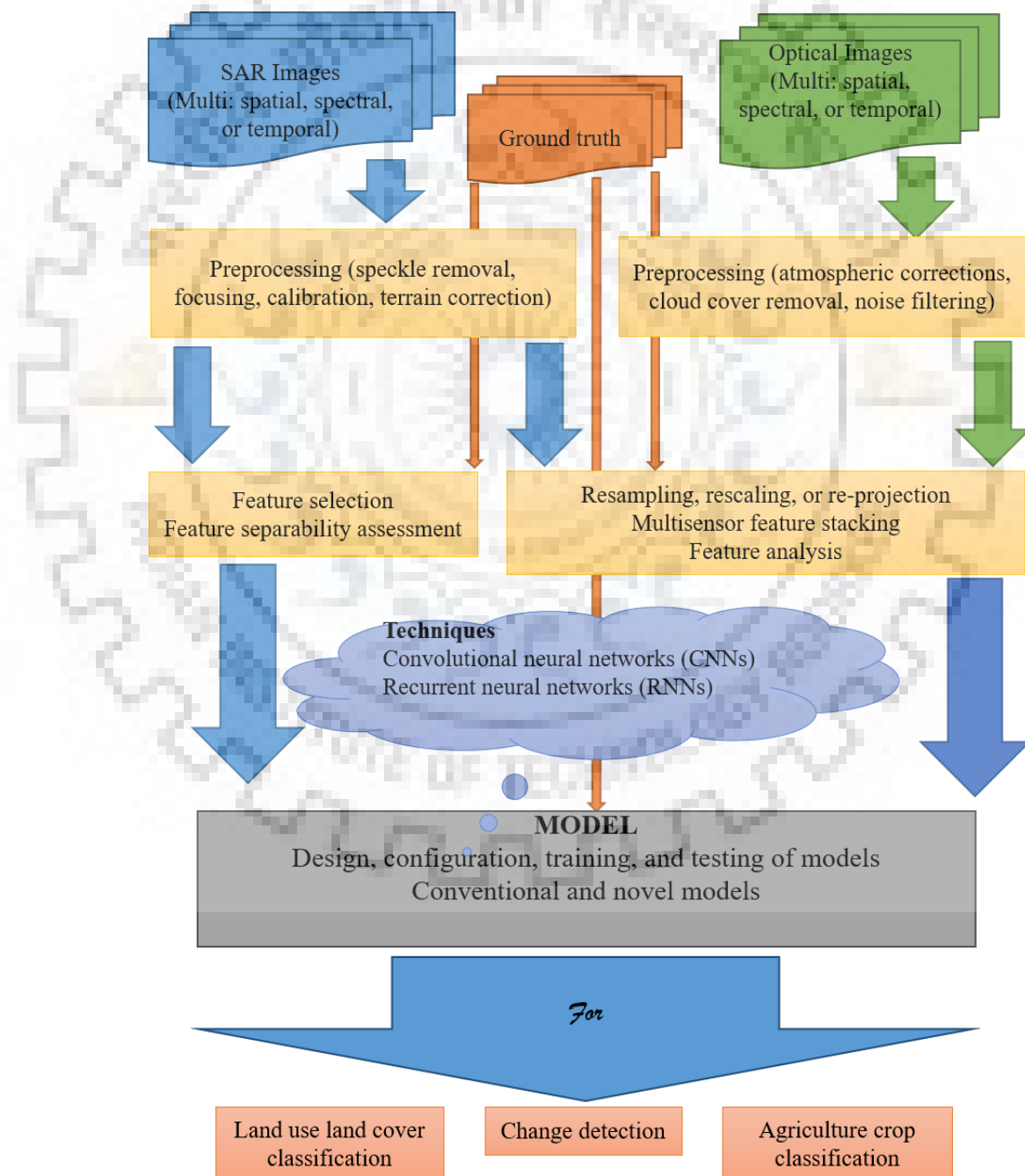
Figure 1.9 Framework of the research

1.8. Organization of the Thesis

The thesis consists of seven chapters. ***Chapter 1*** includes introduction, major challenges, motivation, objective, and tasks presented in the thesis.

Chapter 2 provides brief literature review on the several state-of-the art DNN based techniques reported for various satellite image based applications. The review includes literature on DNN based techniques developed for single sensor, multisensor, and time series satellite images.

The study presented in ***Chapter 3*** involves *a* critical analysis of polarization signatures (PSs) as potential features for PolSAR data based land cover classification, and development of novel approaches which utilize polarization signatures to provide improved classification performance. The study proposed in this chapter is divided into three sub tasks i.e. *sub task 1:* critical analysis of PSs as potential PolSAR features, *sub task 2*: development of an optimal class boundary estimator algorithm to fully exploit PSs, and *sub task 3*: development of CNN model to exploit PSs directly in order to achieve improved land cover classification.

Chapter 4 presents a novel approach to detect human settlements affected by a flood from PolSAR imagery using a novel PSs-CNN based classifier, and image morphology. The object detection algorithm consists of a CNN framework for image classification and a set of morphological operators and *apriori* knowledge for post-classification object detection. The study proposed in this chapter comprises of three stages: —*first*, design and development of PSs based CNN model for extraction of built-up areas from PolSAR image which contains the flooded region, —*second*, development of novel methodology for design of village detection filters based on *apriori* knowledge and image morphology, and —*third*, development of novel methodology for identification and mapping of flood affected villages using image morphology.

Chapter 5 presents study and analysis of conventional and further development of novel multisensor satellite image classification approaches that are based on CNNs. First, the conventional CNN models are analysed and evaluated for multisensor satellite image based land cover classification. Second, two novel multisensor satellite image classification approaches that utilize CNNs in both spatial and spectral domains are proposed and developed. The frameworks employ permuted localized spectral convolutions along with localized spatial convolutions on high dimensional multisensor images. The study establishes the significance of performing permuted localized spectral convolutions in comparison to non-localized or non-permuted localized spectral convolutions. Two models, namely a permuted local spectral convolutional network (Perm-LS-CNN) and a permuted local spectral-spatial convolutional network (Perm-LSS-CNN) are obtained.

Chapter 6 presents a study in which an attempt is made; to analyse the conventional CNN and RNN models, and to develop novel CNN and RNN (in standalone as well as synergistic modes) models for agriculture crop classification with multisensor/multimodal satellite time series imagery. For this purpose, two sub tasks are set in this study. In the first sub task i.e. *sub task 1*, conventional CNN and RNN models are analysed and evaluated for the task of multisensor satellite time series (MSTS) images based crop classification. 1D, 2D, and 3D CNNs, RNNs, and combinations of these are considered for analysis. Then, based on the critical analysis of the working and performance of these conventional models over (MSTS) images based crop classification, 6 novel models are proposed and developed in *sub task 2*.

Finally, ***chapter 7*** concludes the different studies reported in the thesis. The conclusions drawn from each study, contributions made by the studies, and future scopes of the studies are also listed in this chapter.

Chapter 2

BRIEF LITERATURE REVIEW: RECENT TRENDS AND CHALLENGES IN DEEP LEARNING AND DNNS BASED SATELLITE IMAGE PROCESSING

Deep learning and deep neural networks have shown substantial potential as successful satellite image processing techniques. It has been established that DNNs and deep learning (DL) can be extended and generalized to satellite images from natural images [4]. In the recent years, many satellite image based applications such as land cover classification, crop monitoring, building extraction, etc. have been proposed and developed with the help of DNNs and their popularity is still growing [69], [70].

The deep neural network models are high performance, scalable, automatable, and can learn directly from raw data [35]. The deep learning models improve their performance with increase in data whereas other machine learning techniques, which depend on hand-crafted features for learning, are unable to learn beyond a certain amount of data. This means, more the training data, better the deep learning model performs. Another advantage of deep learning is its ability to learn directly from data and independency from hand-crafted features. Traditional machine learning techniques require hand-crafted features for learning. These attribute of deep learning models placed them in prominent position to be used for satellite image processing. This chapter presents a brief review of algorithms developed, challenges faced, and applications proposed in the field of deep learning and DNNs based satellite image processing.

The review involves the contribution and significance of deep learning and DNNs in uni- or multi-modal satellite image processing, single sensor or multi sensor image processing, different satellite image applications, single time stamp or multi-temporal satellite images, and any combination of the above. The review is divided into three categories; -*first:* brief review of recent developments in DNNs based unimodal satellite image processing analysis, and applications, -*second:* brief review of recent developments in DNNs based multisensor satellite image processing, analysis, and applications, and -*third:* brief review of recent developments in DNNs based models for satellite time series image processing, analysis, and applications. Since the studies on DNN based satellite image processing cover loads of fields, the current review is however focussed but not restricted to classification (e.g. land cover classification, crop classification).

2.1. Brief Review of Recent Developments in DNNs based Unimodal Satellite Image Processing

This section presents a brief review of the developments in DNNs based unimodal satellite image processing. Unimodal here refers to a single modality of satellite images i.e. either optical (surface reflectance), PolSAR (backscatter), thermal (brightness temperature) etc. This means that the studies reviewed in this section are developed or evaluated on a single modality of satellite images, may it be optical or PolSAR or any other. Studies on multimodal or multisensor satellite image processing with DNNs are not reviewed in this section. Studies on multimodal or multisensor satellite image processing with DNNs are reviewed in section 2.2 Also, this section does not include studies involving time series satellite images. Such studies are discussed later in section 2.3.

Among other applications, land cover classification is one of the most popular application in the geoscience and satellite image processing community. Numerous techniques have been successfully developed for land cover classification with unimodal satellite images over the years [44]. In context, DNN models have also been successful in satellite image based land cover classification applications.

Satellite SAR images are recently one of the most popular and most widely used modality among all modalities of satellite images. As first attempts in SAR, deep learning based methods have been adopted for a variety of tasks, including land cover or land cover classification [75], [76], object detection [77], parameter inversion [78], despeckling [79], and specific application in InSAR [80].

Land cover classification using PolSAR images is rapidly advancing with the help of deep learning. For land cover classification from SAR and Polarimetric SAR (PolSAR) images, effective feature extraction is essential. Deep learning feature extraction has proved to overcome this challenge, to some degree [75]. Regarding feature extraction, most conventional methods rely on exploring physical scattering properties and texture information in PolSAR images. Compared to conventional methods, deep learning is advantageous in land cover classification due to its capability of automatically learning discriminative features. Moreover, deep learning approaches such as CNNs can effectively extract not only polarimetric characteristics but also spatial patterns of PolSAR images [75]. Deep learning is applied to land cover classification using PolSAR images at first in [81]. They employed a stacked auto encoder (SAE) to automatically learn deep features from PolSAR data and then fed them to a softmax classifier. Remarkable improvements in both classification accuracy and visual effect proved that this method can effectively learn a comprehensive feature representation for classification purposes.

38

Geng et al. in [82] proposed a deep convolutional autoencoder (DCAE) for automatically extracting features and performing classification. The first layer of DCAE are a grey level co-occurrence matrices and Gabor filters. The second layer of the DCAE integrates correlated neighbour pixels to reduce speckle. Following these two hand-crafted layers, a trained SAE, which is similar to [81], is attached for learning more abstract features. The method achieved remarkable classification accuracy when tested on high-resolution single-polarization TerraSAR-X images. Similarly, Qin et al. in [83] applied adaptive boosting of RBMs to PolSAR image classification, stacked sparse AE to PolSAR image classification by taking into account local spatial information is developed in [84], Zhao et al. in [85] proposed a discriminant deep belief networks (DBNs) for land cover classification.

Moreover, taking into account that most current deep learning methods aim at exploiting features either from polarization information or spatial information of PolSAR images, Gao et al. in [86] proposed a dual-branch CNN to learn features from both perspectives for land cover classification.

Different variations of CNNs have been used for land cover classification with satellite PolSAR images. A 6-channel covariance matrix is first extracted in [87] and then, the 6-channel covariance matrix is fed into a trainable CNN model to achieve land cover classification. In [88], the authors focused on the computational efficiency of deep learning methods, proposing the use of lightweight three dimensional CNNs. They showed that classification accuracy comparable to other CNN methods was achievable while significantly reducing the number of learned parameters and therefore gaining computational efficiency. Wang et al. in [89] proposed a fully convolutional network integrated with sparse and low-rank subspace representations for classification of satellite PolSAR images. The knowledge of target scattering mechanism interpretation and polarimetric feature extraction is used to improve land cover classification performance in [90]. He et al. in [91] proposed the combination of features learned from nonlinear manifold embedding and applying a fully convolutional network on input PolSAR images; the final classification was carried out in an ensemble approach by SVM.

The techniques mentioned in the last paragraph are focussed to develop DNN models for either preprocessing SAR images or obtaining desirable features that can be learned with DNNs. Direct utilization of raw SAR or PolSAR data with DNNs could lead to simpler models which can extract abstract features directly from raw PolSAR data based on input-output relationship. In this direction, Zhang et al. in [92] proposed a novel complex-valued CNN specifically designed to process complex values in PolSAR data with CNNs. Other notable complex-valued deep learning approaches for classification using PolSAR images can be found in [93]–[95].

Satellite hyperspectral sensors capture images in hundreds of contiguous narrow spectral bands, producing three-dimensional hyperspectral imagery (HSI). The HIS images provide spectral and spatial information. With the launch of EnMAP, scheduled in 2020, and DESIS, which was planned originally for 2017, high-quality satellite hyperspectral data will be available. Such wealthy spectral information data (with appropriate spatial context) can potentially reveal interesting unknown content within the hyperspectral images. In particular, hyperspectral images have been used widely in a range of practical applications, such as land cover classification, change detection and agriculture monitoring. Indeed, applications of hyperspectral image data have been one of the most active research interests.

Deep learning methods have recently been introduced into hyperspectral image classification, achieving impressive results [96]–[99]. In hyperspectral images, spectral resolution is relatively more important as compared to spatial information. In deep learning (DL) based spectral feature classification methods, DNN models take raw spectral vectors as inputs and output deep spectral features, which are used then for classification. Hu et al. in [100] attempts to carry out HSI classification using 1D-CNN that contains four layers: one convolutional layer, one pooling layer and two fully connected layers. The study in [101] exploits a similar 1D-CNN to classify hyperspectral images but relies on using additional techniques such as batch normalization and dropout. Among the three typical DL models (deep belief networks, CNNs, and stacked autoencoders), stacked autoencoders and deep belief networks take the data represented in a vector form as input, thereby satisfying the need to extract deep spectral feature from spectral vectors. In fact, stacked autoencoders and deep belief networks were introduced prior to CNNs to perform HSI spectral feature classification.

Spectral feature classification methods just take account of spectral information and fail to make use of spatial information. Spatial information (also termed contextual information) helps reveal useful relationships between adjacent pixels. Making full use of spatial information such as shape or size can significantly improve classification accuracy and efficiency. The most significant advantage of 2D-CNNs is that they are suitable to effectively learn features from raw 2D images directly. However, as hyperspectral image data contains more than 100 spectral bands, direct utilization of a 2D-CNNs to extract deep features from raw hyperspectral image data require a mass of learnable kernels. The learnable kernels are hard to train and will increase computational overheads. A randomized principal components analysis (PCA) is employed in [102] to condense the spectral depth of hyperspectral image data in order to improve computation cost. They then applied a 2D-CNN to extract deep features from the reduced spectral space before classification. A similar study is proposed in [103] with the difference in consideration of principal components and window size. They used only 3 principal components and a window

of 42×42 for 2D convolution. Similar studies utilizing feature reduction prior to DL based feature extraction or classification are reported in [104]–[107].

Three-dimensional hyperspectral image simultaneously involves rich spectral information and spatial information. Neither spatial feature classification or spectral feature classification frameworks takes full advantage of this intertwined character of hyperspectral imagery. Deep learning based spectral-spatial classification strategies are more suitable for hyperspectral images and can lead to the improvement of classification accuracy. There are two main strategies to combine the spectral and spatial information present in hyperspectral images; one is to extract spectral and spatial features separately first, and then to combine these features; the other is to extract deep spectral-spatial features from hyperspectral image directly. The former strategy is employed in [103], [106], [108] whereas the later strategy is employed in [96] and [98].

In case of multispectral satellite images or satellite images with low spectral depth, per pixel classification does not yield significant results with deep learning techniques. However, such images, when used in patch mode yield new applications and results. One application of low spectral satellite images with deep learning is scene or patch classification. Scene classification aims to automatically assign a semantic label to each scene image [109]–[111]. Here, a scene image usually refers to a local image patch manually extracted from large-scale high-resolution aerial or satellite images that contain explicit semantic classes (e.g., residential area, commercial area, etc.). Due to high resolutions in such data, different scene images may contain the same types of objects or share similar spatial arrangements between objects. For example, residential and commercial areas may have buildings, roads, and trees, but they belong to two different build-up areas. Indeed, significant variations potentially existing in scene images in the spatial arrangements and structural patterns make scene classification a considerably challenging task [58]. Particular to land cover classification, a patch-based CNN which was more suitable than the pixel-based CNN for medium images is proposed in [112].

In summary, enough evidence suggests the significance of using DL and DNNs for unimodal satellite image processing. Most DL based studies focussed to either PolSAR, hyperspectral, and multispectral satellite images. DNN models have been developed and utilized with PolSAR images for land cover classification, sea ice monitoring, speckle filtering, and change detection applications. On the other hand, spectral and spectral-spatial characteristics based hyperspectral image classification seems to be popular recently. Initially, feature reduction is opted before the DNN model based classification process but later on, hyperspectral images are used directly for classification. 1D-, 2D-, and 3D-CNNs are popularly used with hyperspectral images in various applications including land cover classification. In case of

multispectral satellite images, spatial and temporal characteristics are more focussed and are exploited with the help of 2D-CNNs (spatial), and 3D-CNNs and RNNs (spatial, temporal). Multispectral images are mostly used in scene or patch classification.

However, certain aspects still need attention or are worth analysing. One aspect is, direct exploitation of PolSAR satellite images via DNN models for applications such as land cover classification. In particular, the use of 2D-CNNs to directly extract complex spatial feature from PolSAR data is motivating. More research gaps in DNN based unimodal satellite image processing is summarized in section 1.5.

2.2. Brief Review of Recent Developments in DNNs based Multisensor Satellite Image Processing

DNNs are not restricted to handling unimodal datasets. Instead, exploiting multisensor or multimodal data is one of the advantages of DNNs [60]. It is important to note here that various satellite image applications require to focus on one, or more characteristics (spatial, spectral or temporal) of satellite images. Usually, a single sensor or unimodal satellite images cannot provide high-quality information in all three aspects. Therefore, the use of multisensor or multimodal satellite images for many applications is not uncommon.

This section presents a brief review of the developments in DNNs based multisensor/multimodal satellite image processing and analysis. Multisensor here refers to a set of at least two modalities of satellite images. Multimodal satellite image processing falls under the umbrella of multisensor satellite image processing with image sources from two different modalities. The most popular multi-modal satellite image set is a stack of optical-PolSAR satellite images. Studies on multisensor satellite image processing with DNNs are reviewed in this section with a focus on land cover classification application.

Land cover classification with multisensor satellite images is achieved via two modes. In the first mode, the images from multiple sources are fused together via a fusion process (data fusion), and then the fused image is used as input during the classification process. In the other mode, images from multiple sources are directly used in a stacked manner during the classification process. The development in DNNs based methodologies in both these modes is briefly reviewed here.

Data fusion is one of the fast-moving areas of satellite image processing. Due to recent increases in multisensor data availability, using heterogeneous and big data for earth observation has become more tangible. In this regard, DNNs are natural candidates to tackle the challenges

of modern satellite image data fusion. Pansharpening, feature level, and decision level information fusion are the three major strategies of data fusion.

Pansharpening is the task of improving the spatial resolution of multispectral images by fusing these with high spatial resolution image such as panchromatic image. Traditionally, the field was dominated by works fusing multispectral data with panchromatic bands, but more recently it has been extended to hyperspectral images in [113] and thermal images in [114]. A first attempt in DNN based pansharpening is reported in [115] where the authors used a shallow network to upsample the intensity component obtained after the intensity, hue, and saturation transformation of the red, green, and blue images. Once the multispectral bands have been upsampled with the CNN, a traditional Gram–Schmidt transform is used to perform the pansharpening. The authors simply performed the interpolation operation with a CNN instead of using a nearest neighbour or bicubic interpolation function. Around the same time, authors in [116] used a three hidden layered CNN framework to learn the pansharpening operation end to end. In short, the CNN performs the complete pansharpening process. in this model, authors stacked the panchromatic band with the upsampled spectral bands and then, for each patch, the high resolution values of the central pixel are learned. This method eliminated the element of external intervention in the process and the requirement of alternative methods in the pansharpening process. They used the architecture on GeoEye-1 (panchromatic-multispectral pair), IKONOS-HRV (panchromatic-multispectral pair), and WorldView-2 and achieved satisfactory results. Moreover, the authors in [117] fine-tuned a superresolution CNN model pretrained on natural images for a set of hyperspectral images. They attempted transfer learning [118] between the spectral domains of natural images (three bands, large bandwidths) and the spectral domains of hyperspectral images (many bands, narrow bandwidths).

Most of the current literature dealing with DNNs based satellite image processing, addresses the problem of image classification (e.g. land cover classification). In the case of multisensor or multimodal images, land cover classification is performed by fusing information from multiple sources at either feature level or at decision level. A very simple way of using multisensor images in DNN models for land cover classification or any other application for that matter is to stack or concatenate the multisensor images and process them further as one. The filters learned by the first layer of the DNN, in this case therefore, depends on the order of stack of different sources. Such models have been used in satellite image patch classification. Fully convolutional and de-convolutional models are good examples of such models and have been used in multisensor satellite image classification (patch level) [119]. In contrast to stacking the input images directly, intermediate features (using a feature extraction process) have also been stacked and been used for classification [120]. The models proposed in [120] has been used in

the classification of SAR images via learning through a stack of spatial features. Similarly, in [121], the authors present a hierarchical feature learning process that learns combinations of spatial filters extracted from hyperspectral image bands and digital surface model (DSM) image..

Another series of studies which include an explicit decision level fusion layer between land cover class label images. The studies either perform the decision fusion post classification or the decision fusion process is learned directly. If the classification is performed with two different classifiers one of which is a DNN based model, the different class labels are fused posterior to classification. Two studies are particularly notable in this regard. The authors in [122] fused classification maps obtained from CNNs with classification maps obtained from random forest (RF). The RF classifier is trained on handcrafted features. The two maps are fused by multiplication of the posterior probabilities. An edge-sensitive conditional random field is also learned on top to improve the quality of the final labelling. Whereas the authors in [123] learned an ensemble of CNNs and then averaged their predictions. Their proposed processing framework has two main streams: -*first,* processing the colour infrared data from the Vaihingen dataset, and –*second,* processing the digital surface model. They trained an ensemble of CNNs and further used them as inputs for each layer of the primary model's activation maps. They obtained a series of land cover maps to nourish the ensemble. This strategy improved the performances by considering classifiers issued from different data sources and levels of abstraction. This model has the advantage of being entirely learned in an end-to-end fashion. However, it incurs a heavy computational load and involves a complex architecture with numerous fusion layers and skip connections. Data fusion is not only about fusing image data with the same viewpoint. Multimodal satellite image data that exceed these restrictive boundaries and approaches to tackle new, exciting problems with satellite image processing are beginning to appear in the literature.

In summary, DNNs based multisensor satellite image processing, analysis, and applications have significant potential and future. Notably, multisensor satellite image based land cover classification with DNNs provides higher accuracy, less fine-tuning, high dimensional data operability, and scalability. A model developed for a particular multisensor satellite image set can be used with other multisensor satellite image sets with slight modifications. Although, data fusion approach for multisensor satellite image applications seems to be popular right now but direct utilization of multisensor satellite images with DNNs is gaining attention. Few research gaps identified in multisensor satellite image processing with DNNs are discussed briefly in section 1.5.

2.3. Brief Review of Recent Developments in DNNs based Satellite Time Series Image Processing

This section presents a brief review of the developments in DNNs based satellite time series image processing techniques. Satellite time series images refer to a set of satellite images collected over time observing the same area of interest with the same sensor. If more than one sensor is considered, then the sensors may belong to same modality (unimodal) or different ones (multimodal). The review includes studies on both, unimodal and multisensor satellite time series images. However, in this section, the review is presented in the order of DNN technologies used in satellite time series image processing. The popular techniques include CNNs, RNNs and combined CNNs-RNNs.

CNNs are the most popular DNN technology in satellite time series image processing. In the recent years, numerous techniques have been developed for satellite time series image processing using 1D, 2D, and 3D CNNs. 1D CNNs are first utilized for Landsat-8 satellite time series image processing in [124]. They used 1D-CNNs with multilayer perceptron inspired from [125] to perform land cover classification. Zhong et al. in [126] developed two types of deep learning models; first, a 1D CNN based satellite time series crop classification model with inception blocks within the classification framework and second, a long short term memory (LSTM) RNN based crop classification model. The 1D-CNN with inception blocks model performed well in contrast to basic LSTM- RNN model for satellite time series image based crop classification. Pelletier et al. in [73] proposed a temporal CNN model to perform satellite time series image based crop classification. The model focuses the significance of temporal convolution in time series image classification. They also evaluated the effects of appending hand crafted features such as vegetation indices into the classification model. Their model performed satisfactorily over random forests and RNNs. Further, 3D-CNNs have also been utilized for satellite time series image processing in many studies. 3D CNNs have the ability to capture spectral, spatial, and temporal features simultaneously whereas 1D and 2D CNNs lack such ability. A satellite multi-temporal images based crop classification model is proposed and developed in [127] which uses 3D-CNNs as spatiotemporal feature extractors. It is validated from the study that 3D-CNNs are better spatiotemporal feature extractors than 2D or 1D CNNs for satellite time series images. A 3D CNN-transformer hybrid approach for crop classification using multitemporal Landsat-8-Sentinel-2 images is developed in [128]. The transformer architecture utilizes multi-head self-attention module to represent the sequence pattern which improves the classification performance in contrast to conventional RNN based crop classification modules.

RNNs are also popular DNN architectures to utilized with satellite time series images. Along with CNNs, RNNs have also gained interest in satellite time series image applications. The first study which utilized RNNs with satellite time series images is proposed in 2016 [129]. They developed a transferable change rule algorithm based on LSTM-RNNs for detecting change in land cover. Their comparison with SVM-CNN and decision tree signified the better performance of the proposed model. Later, a LSTM-RNN based model is developed in [74] for crop classification. The model is evaluated on Sentinel-2 multispectral time series images and performed well in comparison to CNNs and SVMs. A similar model is proposed in [130] for pixel as well as object based land cover classification. This model is evaluated over two different multispectral datasets and results indicated improved and stable performance. Moreover, RNNs have also been utilized for PolSAR time series image based crop classification in [131] and [132]. Both, LSTMs and gradient recurrent unit (GRU) based RNN models have been developed and tested with Sentinel-1 time series images. Both studies indicate to a slightly better performance of GRU over LSTMs. Sun et al. in [133] used ancillary time series cropland data layer along with Landsat-8 time series images and further used LSTM-RNN model over the combined dataset to perform land cover classification. Another study which uses a patch or window of pixel instead of per-pixel or object based in classification process is proposed in [134]. The study is developed on Landsat-8 time series images and LSTM-RNN architecture is used. The patch based land cover classification model benefits from the contextual information present in the patch which otherwise is absent in pixel based classifications.

CNNs and RNNs have also been used synergistically and successfully in satellite time series image processing. A 2D-CNN-RNN architecture is developed in [135] to perform data fusion with multimodal, multiscale, and multitemporal satellite images. The satellite time series images are Sentinel-2 whereas a very high spatial resolution SPOT-6/7 image is also used. In another study, a temporal attention network is developed for multisensor and multitemporal satellite time series image based crop classification. Attention mechanism is a brain mechanism of human signal processing. The study utilizes attention network in the temporal dimension to reinforce the phenological differences in crops and restraining their phonological similarity. Further, in [136], CNNs are used with RNNs in a parallel architecture to extract spectral-spatial-temporal features from a combination of MODIS time series and NAIP high spatial resolution dataset. Finally, a SVM and multilayer perceptron models are used to perform crop classification. CNN-RNN models have also been used in change detection applications. In [137], authors used a recurrent convolutional cascade model for learning spectral-spatial and temporal features from Landsat -7 time series multispectral images. The learned features are then used to detect changes with a binary classification module.

In summary, three major DNN model categories i.e. CNNs, RNNs, and CNNs-RNNs have been successful in analysing and processing satellite time series images for various applications. The applications include land cover classification, change detection and most importantly, agriculture crop classification. Although CNNs and RNNs have performed well with satellite time series images in standalone modes, synergistic utilization of both models either in parallel or in cascade fashion, results in superior performance. The use of CNN-RNN architecture in feature extraction and classification with multisensor satellite time series images for crop classification is very motivating. Research gaps in DNN based multisensor satellite time series image processing are discussed briefly in section 1.5.

2.4. Conclusions

This chapter presented a brief review on developments in DNN models and methodologies for satellite image processing, analysis, and applications. The review included unimodal satellite images, multisensor satellite images, and satellite time series images. In context to applications of satellite images, focus of this review is kept to classification majorly (for example, land cover classification, crop classification). However, other applications are also touched upon. The review consolidates to certain observations.

- Most of the novel DNN models and methodologies have been explored for SAR, hyperspectral, and multispectral satellite images.
- Spatial or contextual information present in satellite images is given preference in contrast to spectral information (in context to multisensor satellite images).
- Both parallel and cascade DNNs architectures have been developed. Parallel architectures extract spatial and spectral information simultaneously with two different DNN modules whereas the cascade architectures used the same module to extract the spatial and spectral information (one after the other). Critical analysis of these architectures is worth investigating and interesting.
- CNNs are more popular than other DNN models while handling spectral or spatial features of satellite images are whereas RNNs and 1D-CNNs are more popular while handling temporal features of satellite images. However, integration of CNN and RNN models for time series satellite image processing might yield high performance and consistency and also could lead to development of a general framework.
- The main focus of most studies is on classification tasks for example, land cover classification, scene classification or, object detection. However, several other applications are worth mentioning including fusion, segmentation, change detection, and registration.

Prominent research gaps are identified and are briefly listed in chapter 1 (refer section 1.5).

Chapter 3

CNNs and Polarization Signatures Based Novel Approach for Land Cover Classification with PolSAR Satellite Images

Use of CNNs with PolSAR satellite images for various earth observation applications is increasing at pace. One application where the use of CNNs has increased rapidly is land cover classification. However, the focus of new CNN based models for PolSAR image based land cover classification is to achieve higher classification performance in mixed land cover class scenarios. This chapter proposes a novel approach based on CNNs which exploits PolSAR images to improve classification performance in mixed class scenarios. This chapter highlights the challenges while extracting information from PolSAR images for land cover classification. One major challenge while extracting land cover information from PolSAR satellite images is to segregate or classify land covers efficiently in a mixed land cover scenario. Therefore, this chapter presents two different novel approaches, *-first* optimal class boundary estimator based decision tree classifier, and *–second* CNN model based classifier to extract information form PolSAR images for improved land cover classification in a mixed land cover scenario.

3.1. Introduction

Use of polarimetry for radar applications has a long history. PolSAR target decomposition based techniques are used by various researchers for many earth observation applications such as land cover classification [39], [138]–[140]. A typical land cover scene includes both, natural (e.g. forests, water bodies) and man-made (e.g. buildings, bridges, roads) targets. These land covers show complex scattering response and poses a challenge in classification using satellite PolSAR data. Target decomposition based techniques aims at analysing the received scattering matrix from polarimetric data to extract information about the scattering processes exhibited by various land covers. These techniques compute particular features (e.g. decomposition features) from the backscattering response of land covers (recorded by the PolSAR sensor) and further use these features to detect, identify or, classify land covers. An extensive study on the performance of various decomposition based features in land cover classification is done in [141]. They observed an absence of dominant decomposition based features for different land covers in mixed land cover scenario. Mixed class scenarios refer to landscape (study regions) where land cover classes considered for classification are small, unorganized, and are scattered randomly. Due to this

landscape two or more different land covers appear similar in a satellite image. For example, built-up areas having building structures close to each other and are randomly oriented (with respect to SAR sensor flight path), might show volume scattering along with double bounce scattering thus behaving as mixed class. Researchers are also developing new multi-component decomposition and factorization based features that are able to segregate land covers more efficiently but the challenge still needs more attention [42], [142]–[144] Impact of various target decomposition base features on land cover/use classification performance is analysed in [145]. They merged land cover/use classes that are not separable with given dataset and features. For example, agriculture and bare soil are combined and treated as one class. Similarly, rural settlements and dense tall vegetation are combined and treated as one class. This indicates to the challenge to segregation various land covers in a mixed land cover scenario. An example of mixed land cover scenario is shown in Figure 3.1. For example, in Figure 3.1(b), marker A indicates rural settlements (built-up area) and marker B indicates dense tall vegetation patches. Observing the enlarged segments in Figure 3.1(c), it is evident that both the land cover classes are showing similar backscattering behaviour (high backscattering response in HV channel) and [145] found similar observation. Literature indicates few studies that aim to differentiate class pairs such as bare soli-agriculture or dense tall vegetation- rural settlements. Attempts are made in [42], [146], [147] and more recently in [42], [148] to segregate oriented built-up areas from tall vegetation using different decomposition and factorization features. However, the challenge still exists and new features needs to be explored which might have potential to segregate these class pairs and provide good land cover classification results.

In the context, PolSAR polarization signatures can be potential candidates. PolSAR polarization signatures (PSs) can characterize the scattering behaviour of target under study or in this case land cover under observation effectively and therefore can be considered for land cover classification applications. These signatures are 2D representations of scattering characteristics of a target. A method for generation of PSs with different polarization bases is proposed in [149]. They reported better classification accuracy than object based supervised Wishart classification method. Similar work proposed in [150] compared observed PSs to canonical target or reference PSs based on norm and further, a pattern recognition using Kohonen networks was performed for LC classification in [150]. Results show that PSs not only allow to identify coherent scatters but also scattering mechanisms. Till date, few approaches have been developed for analysing PSs for land cover (LC) classification purposes [151]–[153]. Therefore, PSs are considered here to be analysed as potential PolSAR features for land cover classification in mixed land cover

scenario. In summary, this chapter presents a study in which PolSAR PSs are considered as potential PolSAR features for land cover classification.

However, the challenge with exploiting PSs for land cover analysis is its contextual nature. It means that PSs pixels are related with its neighbourhood. But, the current popular techniques that directly exploit PSs such as the normalized signature correlation mapper (NSCM) and normalized Euclidean distance (NED) based classifiers requires the transformation of the two dimensional PSs into a one dimensional vector before classification. This results in a loss of contextual information present in the PSs. There is a need to develop classifiers which can exploit the contextual information present in the PSs. In the direction of novel classifiers, authors in [146] and [148] utilized deep learning techniques for land cover (LC) classification with the help of synthetic databases for learning. Their attempt is motivating and points to the potential of deep learning techniques in extraction of information form PolSAR images. These studies motivate to explore the use of deep learning and DNNs with PolSAR data for various applications. The study presented in this chapter to propose a CNN based approach for PolSAR PSs based land cover classification.
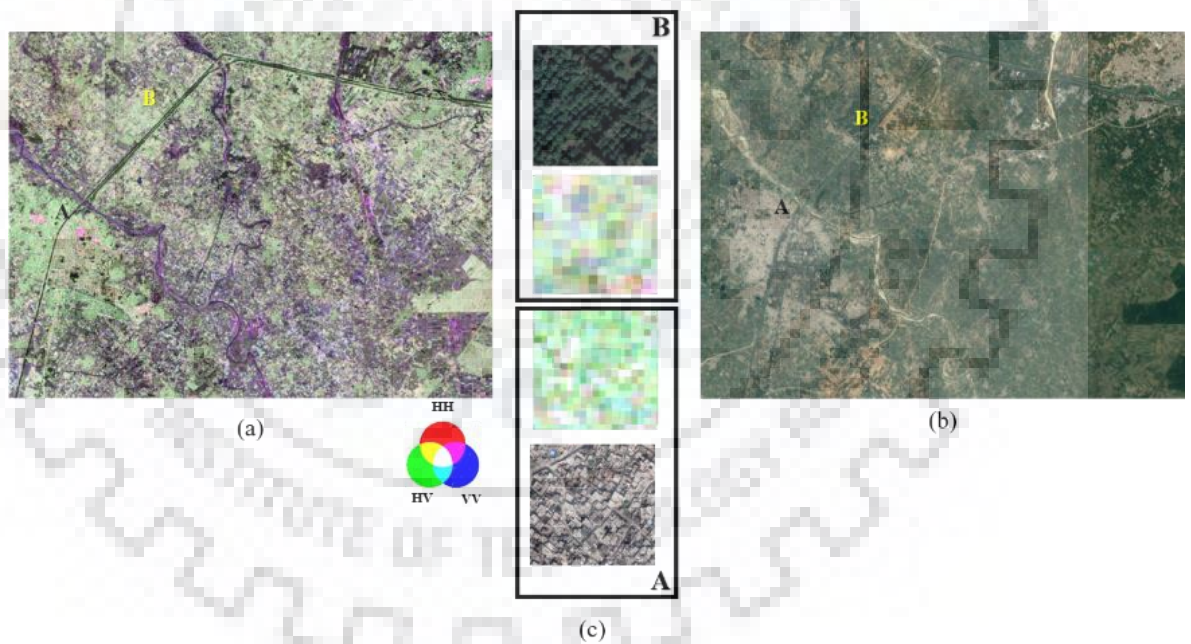
Figure 3.1 (a) False colour composite generated from PALSAR-2 data. Colour scheme: red = HH, green = HV, and blue = VV. (b) Panchromatic image collected from Google Earth dated Feb 19, 2015. Symbol 'A' marks built-up area in both the images and symbol 'B' marks the dense vegetation area in both the images. (c) The enlarged segments are shown in middle in the outlined boxes. The built-up area is showing volume scattering (high backscatter response in HV, refer) along with double bounce scattering.

The study presented in this chapter is divided into three sub tasks i.e. *sub task 1*: critical analysis of PSs as potential PolSAR features, *sub task 2*: development of an optimal class boundary estimator algorithm to fully exploit PSs, and *sub task 3*: development of CNN model to directly exploit PSs.

The first sub-task of the study covers the introduction to polarization signatures and further, the computation of polarization signatures based features from fully polarimetric SAR data. A land cover (e.g. bare soil) behaves as a complex mixture of different canonical scatterers or canonical targets and therefore its PS is difficult to interpret. Therefore, at first, there is a need to find correlations between polarimetric/ PSs of land covers and PSs of canonical targets. These correlations indicate the degree of similarity between the two scatterers. Further, the PS correlation features or PSCFs needs to be critically analysed as potential features for PolSAR image based LC classification. Section 3.4 cover the details on the computation of PSCFs and their critical analysis as significant PolSAR features for LC classification.

. In the second sub task of the study, a novel adaptive and optimal class boundary estimator is developed. Image statistics are used to obtain class boundary values which make the approach robust and adaptive. Genetic algorithms are used to optimize the class boundary values to achieve maximum overall classification accuracy. A set of optimal class boundaries are obtained using this estimator, and a decision tree (DT) classifier is realized using these class boundaries. Section 3.5 presents the development and implementation of the adaptive and optimal DT classification technique.

At last, in the third sub task of the study, a second novel classification approach namely PSs-CNN which utilizes CNNs as classifier directly with polarization signatures to perform land cover classification is developed. Since, the PSs are two dimensional representation of target backscatter response, they can be treated as images of target response and thus can intuitively be used in a 2D-CNN classification framework. The 2D-CNNs are able to captures the spatial context present in a pixel PSs which helps in the classification of the pixel. Section 3.6 presents the PSs and CNNs based novel classification technique.

Remaining of the chapter is outlined as follows. Section 3.2 covers the theoretical background required to understand the study. Section 3.3 provides description about the experiment setup which includes; the study areas and datasets considered in the study, and the ancillary information collected for algorithm development and training. Section 3.7 concludes the study.

3.2. Theoretical Background

3.2.1. Polarization signatures

Polarization signature (PS) is a two dimensional representation of backscattering behaviour of a target (in this case, a land cover). In PS representation, x-axis and y-axis represent ellipticity angle and orientation angle respectively and z-axis represent received backscattered power coefficient ($\sigma°$). Orientation or tilt angle (ψ) varies from $-90°$ to $90°$ and ellipticity angle (χ) varies from $-45°$ to $45°$. The signatures are computed using equation (3.1) [154], [155].

$$\sigma\left(\chi_i\psi_i\chi_j\psi_j\right) = \frac{4\pi}{k^2}\begin{pmatrix} 1 \\ \cos 2\chi_i \cos 2\psi_i \\ \cos 2\chi_i \sin 2\psi_i \\ \sin 2\chi_i \end{pmatrix}(\mathbf{K})\begin{pmatrix} 1 \\ \cos 2\chi_j \cos 2\psi_j \\ \cos 2\chi_j \sin 2\psi_j \\ \sin 2\chi_j \end{pmatrix} \tag{3.1}$$

Where σ represents backscattering coefficient, suffix i and j represent transmit and received combinations respectively. $\mathbf{K}$ is the Kennaugh matrix, which provides received power, given the polarization characteristics of the receiving antenna [155]. It is also known as the Stokes scattering matrix. Computation of **K matrix** parameters is provided in Appendix-A. k is the propagation constant. Co-polarized signature is obtained by transmit and received combination of $\psi_i = \psi_j$ and $\chi_i = \chi_j$, and cross polarized signature of $\psi_i = 90° + \psi_j$, and $\chi_i = -\chi_j$. Ellipticity angle defines the polarization behaviour (linear, circular or elliptical polarization), while orientation or tilt angle in combination with ellipticity angle, defines polarization states, for example, horizontal or vertical polarization [156]. For the sake of visualization, co-polarized PS of an arbitrary land cover is shown in Figure 3.2.
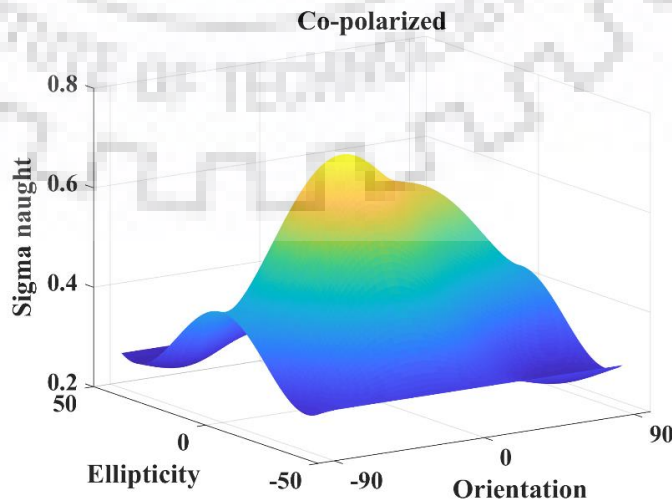
Figure 3.2 Co-polarized polarization signature of an arbitrary land cover.

53

3.2.2. Convolutional neural networks

CNNs are a popular class of deep feedforward artificial neural networks proposed in various studies [157]–[160]. Building blocks of a conventional CNN are: a convolutional layer, a pooling layer, a dropout layer, a fully connected layer, and a loss layer. A generic arrangement of the layers is discussed in [161] and [162].

Imagine a CNN with M layers, the output state vector of the mth layer is denoted by $\boldsymbol{O}^m$, where $m \in \{1, 2...M\}$ and, $\boldsymbol{O}^0$ represents the input data. The input layer is then connected to a convolutional layer. In the convolutional layer, a 1D, 2D or 3D convolution is performed with kernels. For the sake of understanding, a 2D convolution is explained here. The convolution plus pooling in the mth layer is represented in equation (3.2).

$$\boldsymbol{O}^m = \text{pool}(g(\boldsymbol{O}^{m-1} \otimes \mathbf{W}_m + \boldsymbol{b}_m)) \tag{3.2}$$

where, $\otimes$ indicates convolution operation and *pool* denotes a spatial aggregation, as defined in [163]. The matrix $\mathbf{W}_m$, encodes the weights of the links connecting the $(m-1)^{\text{th}}$ convolutional layer to the mth layer. The vector $\boldsymbol{b}_m$ denotes a bias term. Subsequently, a non-linear pointwise activation operation $g(.)$ is performed. Finally, a pooling layer is used which helps in selection of dominant features.

Several convolutional and pooling layers can be stacked to form a hierarchical feature extraction architecture. The resultant features are then combined to form a one-dimensional feature vector to be taken as input by the fully connected layer. The fully connected layer acts as a typical multi perceptron layer. The last layer is usually a linear (for regression scenarios) or softmax (for classification scenarios) layer, with the number of neurons matching the number of output variables or classes desired. This layer also specifies how training penalizes the deviation between the predicted and true labels and is normally the final layer.

The weights, $\{\mathbf{W}_1, \mathbf{W}_2, ... \mathbf{W}_M\}$ and the biases, $\{\boldsymbol{b}_1, \boldsymbol{b}_2, ... \boldsymbol{b}_M\}$ of the CNN compose the model parameters, which are iteratively and jointly optimized via overall accuracy maximization over the training set. Similar networks can be realized for 1D or 3D CNNs accordingly.

The whole network still expresses a single differentiable score function from the raw image pixels on one end to class scores at the other. 2D CNN architectures make the explicit assumption that the inputs are images. This allows the user to encode certain properties into the architecture. These make the following functions more efficient to implement and it also vastly reduce the amount of learnable parameters in the network. This property is explored and utilized

for PSs in this study. CNNs with deep learning have found their application in PolSAR data as well [58].

3.3. Experiment Setup

3.3.1. Study area

The area selected for study includes Roorkee city at the foothills of the great Himalayas of the northern India and its neighbouring region. The area extends from $29°57'4''N$, $77°51'54''E$ to $29°48'10''N$, $78°4'5''E$. The area covers about 400 km^2 of landscape.   The study area includes rivers and water canals, built-up areas include Roorkee city and villages nearby. Built-up areas are both sparse and dense. Structures in built-up areas are in patterns or at random. Area also includes sparse and dense tree plantations, cultivated and fallow agricultural lands, dry and water logged patches of land. Figure 3.1(b) depicts the Google Earth imagery of the study area. This topography makes the study area a mixed land cover scenario in terms of land cover classification.

3.3.2. Dataset

The dataset used for study is the ALOS PALSAR-2 L band synthetic aperture radar (SAR) data. PALSAR-2 is a polarimetric SAR sensor on-board the ALOS-2 spaceborne satellite. SAR data collected by PALSAR-2 is fully polarimetric with information in four polarization states i.e. HH, HV, VH, and VV. PALSAR-2 sensor and product specifications are listed in Table 1.8. The data acquisition date is March 13, 2015. The data is provided by JAXA, Japan for research purposes. A subset of the whole image, nearly 400 km^2 is considered for algorithm development. ALOS PALSAR-2 data has been reported to be utilized successfully for many land cover applications such as geology mapping [164], land deformation monitoring [165], and mostly in land cover classification [166], [167].

3.3.3. Ground truth

Ground truth sample points are collected for training and validation. Five land cover classes are considered namely bare soil (BS), tall vegetation (TV), short vegetation (SV), built-up (BU), and water. Ground truth class sample points are measured directly on the terrain based on visual inspection. Approximately 1200 data points are collected for both the study areas. Details of the ground truth collected for both study areas are shown in Table 3.1. Also, Figure 3.3 depicts samples of ground truth collected during the study.

Table 3.1 Summary of ground truth sample points collected for study.

| Land Cover | Roorkee | |
	Training and validation	Testing
BS	200	40
TV	200	40
SV	200	40
Built-up	200	40
Water	200	40
Total	1000	200
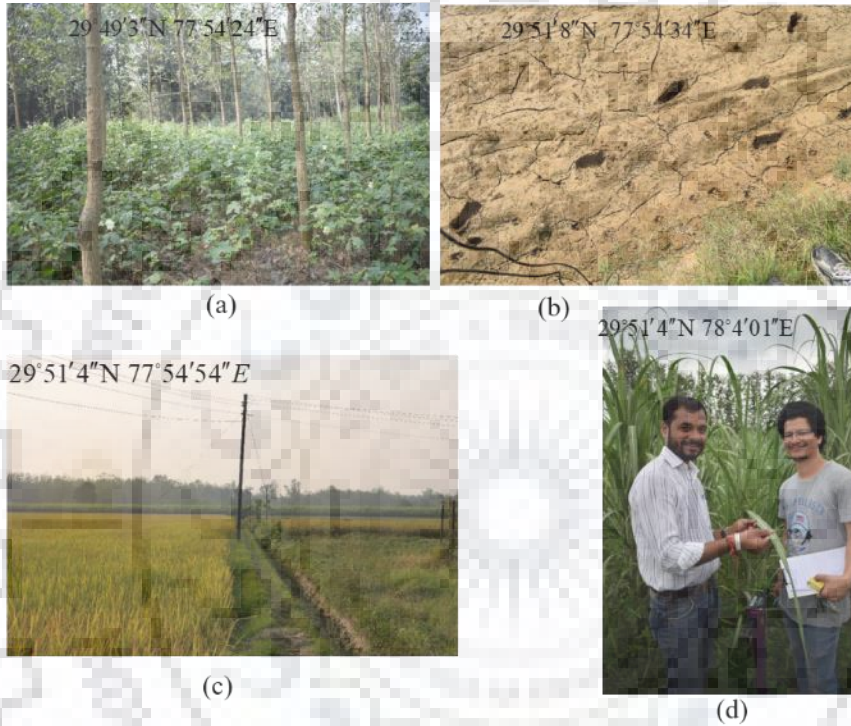
Figure 3.3 Sample images of different land covers collected during ground truth survey. (a) Tall vegetation, (b) bare soil, (c) agriculture cropland, and (d) sugarcane field.

3.3.4. Preprocessing

The PALSAR-2 data considered in the study is a single look complex (SLC) data. Preprocessing is required before using the data for the study. The steps involved in preprocessing are similar to as discussed in [168]. A Gamma MAP filter of window size 3×3 for the purpose of speckle removal is used as this filter retains polarimetric signature characteristics of SAR image [169], and [170]. This filter also removes speckle noise from gamma distributed scenes such as forested areas and agricultural lands whereas keeping the original pixel values for non-gamma distributed scenes such as water. The window size is finalized on the basis of spatial resolution of input image and application. Geocoding is done using SRTM 3 arc-second global DEM model. Other tasks like radiometric calibration and terrain correction are performed similar to [168]. The

output of this step is a pre-processed backscattering or S-matrix. Four features namely HH, HV, VH, and VV are obtained. Since, HV and VH are similar due to the reciprocity theorem, only HH, HV, and VV are used further in the current study.

Figure 3.1(a) show a false colour composite image of the pre-processed Roorkee study area. It is observed from the false colour composite image that some parts of built-up areas are showing volume scattering (represented by green colour) behaviour similar to dense vegetated areas such as tree plantations. For example, in Figure 3.1(a), built-up area (marked as A) is showing backscatter behaviour similar to backscatter behaviour of tree plantations (marked as B). It can be further observed from enlarged patches shown in Figure 3.1(c). It is because of the randomly oriented, densely packed and non-similar structures constitute the built-up area and hence is showing volume scattering as dominant scattering. The main objective of the current study is to develop features and classification techniques that can overcome this challenge and achieve higher classification performances.

3.4. Sub Task 1: Polarization Signatures: Critical Analysis as Potential PolSAR Features for LC Classification

This section discusses the PSs based feature computation process and its utilization for land cover (LC) classification. Section 3.4 covers the sub task 1 of the study. Flowchart shown in Figure 3.4 summarizes the process of, computation of PSs, and its critical analysis as potential PolSAR features.

3.4.1. Polarization signatures computation: Observed and canonical

Due to complex geometrical structure, natural targets (land covers) show complex scattering response [171], [172]. Scattering responses from natural targets are represented as a combination of simple point targets scattering responses which are well defined [173]. In this study, PSs from canonical targets namely *Flat Plate (FP)*, *horizontal dipole (HD), vertical dipole (VD),* and *dihedral (Di)* are considered as canonical targets. Geometrical structures, scattering matrices, and PSs of these canonical targets are shown in Figure 3.5. The PSs of these canonical targets are further utilized in representing PSs of natural targets (observed or image-pixel PSs). Further details about canonical targets are provided in [174]. PSs, both observed (image pixel) and canonical target are obtained using equation (3.1). The PSs are obtained for ellipticity angle χ ranging from -45° to $+45^o$ and orientation angle ψ ranging from -90° to $+90^o$ with a step size of 5° in each direction. Each obtained PS is a 18×37 two dimensional matrix.
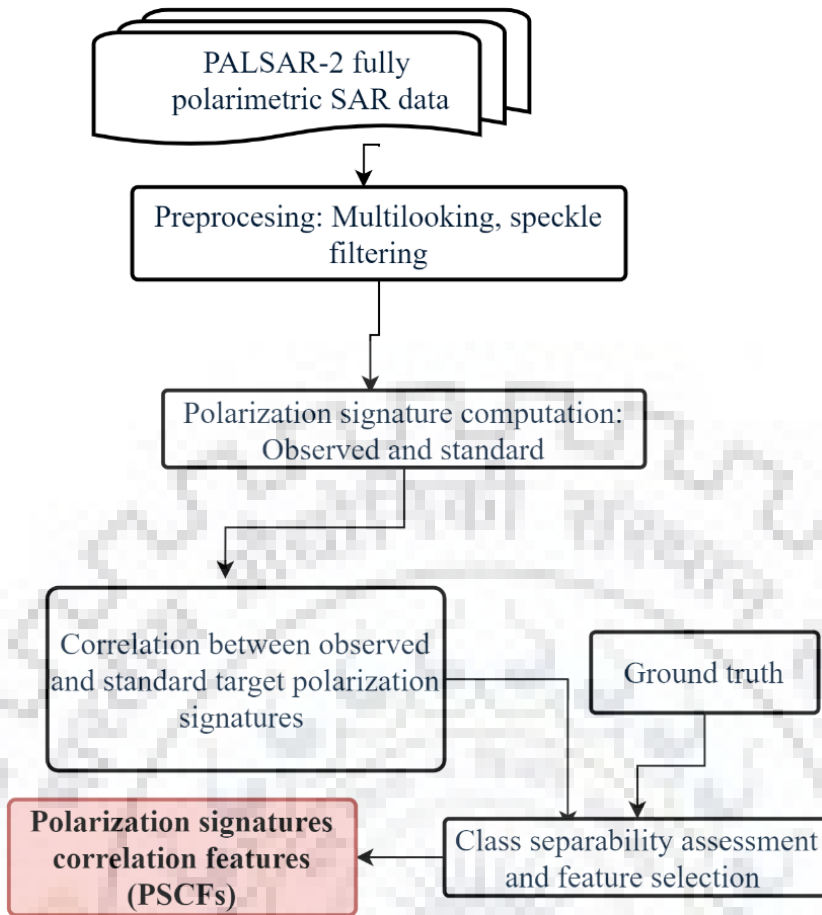
Figure 3.4 Flowchart showing polarization signature correlation features (PSCFs) computation using fully polarimetric SAR data.

$$S = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad S = \begin{bmatrix} \cos{^\wedge}2\emptyset & \frac{1}{2}sin2\emptyset \\ \frac{1}{2}sin2\emptyset & \sin{^\wedge}2\emptyset \end{bmatrix} \qquad S = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$
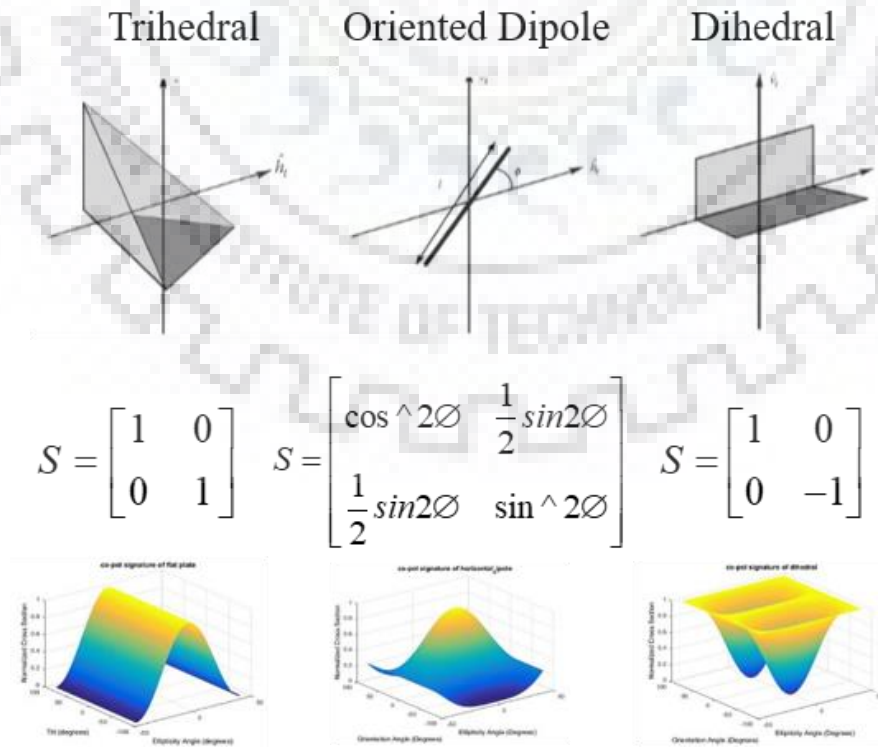
Figure 3.5 Geometrical structures, scattering matrices, and co-polarization signatures of different canonical targets.

58

3.4.2. *Polarization signature correlation features or PSCFs*

A land cover (e.g. bare soil) behaves as a complex mixture of different canonical scatterers or canonical targets and therefore its PS is difficult to interpret. Therefore, at first, there is a need to find correlations between polarimetric/ PSs of land covers and PSs of canonical targets. These correlations indicate the degree of similarity between the two scatterers.

Correlation coefficient is a value, which varies from +1 and −1 and it is calculated to represent the linear inter-dependence of two variables or sets, of data. For two variables, say x and y, the correlation coefficient is computed using equation (3.3).

$$CC = \frac{S_{xy}}{S_x S_y} \tag{3.3}$$

In this section, correlation is computed between each image pixel-canonical target PS pair. Here, x is the observed (image pixel) PS, y is the canonical target PS, S_x is standard deviation in x, S_y is standard deviation in y, S_{xy} is co-variance between x and y, and CC is correlation coefficient between x and y.

The following pairs are considered between which correlation is computed.

- Co-polarized (pixel-dihedral) or Corr_co_Di

The term *pixel* represents observed co-polarization signature from a pixel in image, and term *dihedral* corresponds to co-polarization signature of dihedral canonical target. *Co-polarized (pixel-dihedral)* means correlation coefficient between observed co-polarization signature and dihedral co-polarization signature. Similarly, other pairs are represented as

- Co-polarized (pixel-flat plate) or Corr_co_FP
- Co-polarized (pixel-horizontal dipole) or Corr_co_HD
- Co-polarized (pixel-vertical dipole) or Corr_co_VD
- Cross-polarized (pixel-dihedral) or Corr_cross_Di
- Cross-polarized (pixel-flat plate) or Corr_cross_FP
- Cross-polarized (pixel-horizontal dipole) or Corr_cross_HD
- Cross-polarized (pixel-vertical dipole) or Corr_cross_VD

In total, 8 such pairs are obtained in the process. These features are termed as polarization signature correlation features or PSCFs. These PSCFs are potential features for LC classification. PSCF computation process followed in section 3.4.2 is summarized in Figure 3.6 for the sake of clarity
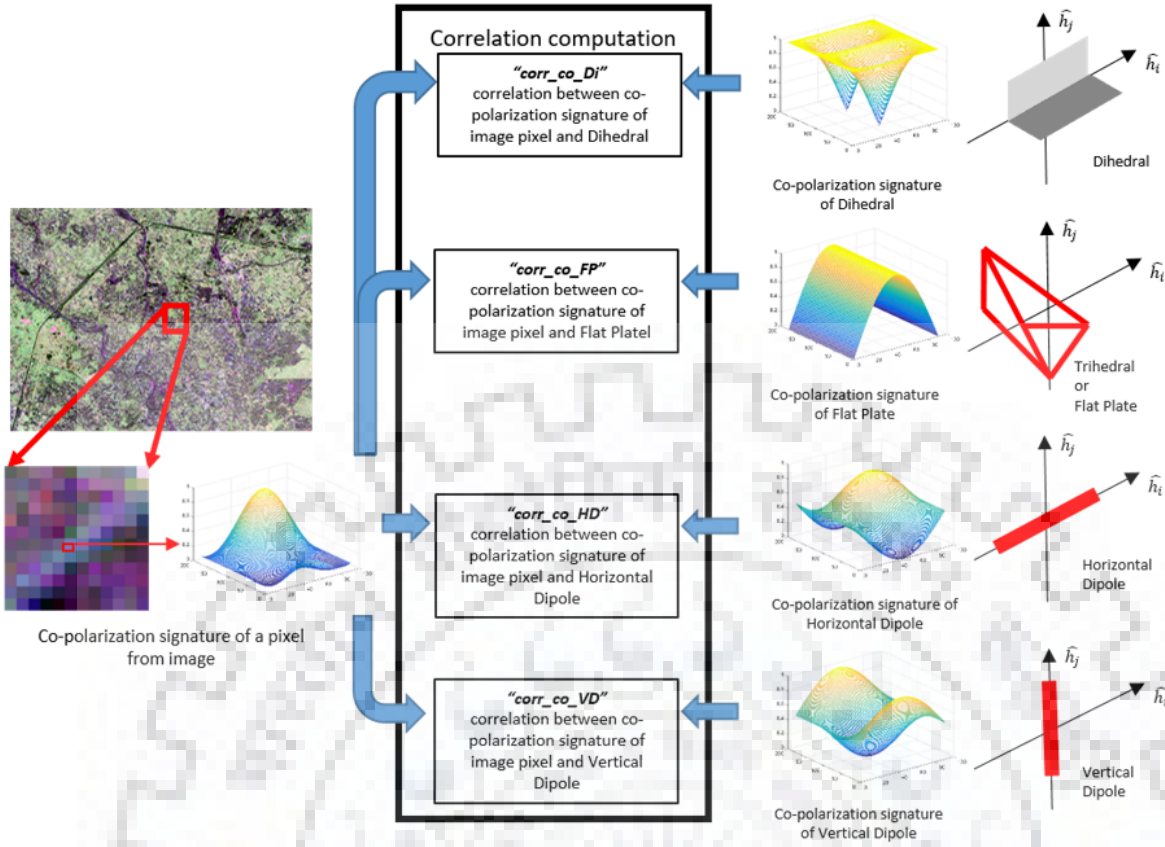
Figure 3.6 Figure illustrates the computation of polarization signature correlation features (PSCFs) between observed (image pixel) and canonical targets PSs.

A critical analysis is essential at this stage in order to identify the best feature for segregating a particular land cover class. For the same, a separability index based feature selection process is carried out in section 3.4.3.

3.4.3. Assessment and critical analysis of PSCFs

In this section, each PSCF obtained in section 3.4.2 are analysed and discussed. Class separation for each PSCF is evaluated by computation of class separability index (SI) proposed in [175]. Nomenclature used in section 3.4.2 is used in this section also.

The separability measure is computed using equation 3.4. For a given feature, $SI_{i,j}$ is the measure of separability between class i and j. μ_i and μ_j are mean values for class i and j respectively, and σ_i, σ_j are standard deviation for class i and j respectively.

$$SI_{i,j} = \frac{|\mu_i - \mu_j|}{|\sigma_i + \sigma_j|}$$

(3.4)

In order to measure SI between different land cover/class pairs, correlation values are collected for all the ROI points (for Roorkee study area) mentioned in section 3.3.3. The SI values obtained for all the PSCFs as features are provided in Table 3.2.

Separability assessment is also conducted for decomposition features obtained using Cloude decomposition [139], modified Yamaguchi four component decomposition or m-Y4D [176], generalized Freeman Durden decomposition or GFD [177], and H-A-Alpha quad pol decomposition or H-A-Alpha [178] methods. SI values for class pairs are shown in Table 3.3.

A comparison between the PSCF features and the considered decomposition features on the basis of SI value for selected class pairs is done. The selected class pairs are, TV to built-up, TV to SV, and BS to Water. These three class pairs are difficult to segregate in a mixed class scenario and therefore are considered for comparison. From Table 3.2 and Table 3.3, it is evident that for TV to built-up class pair, corr_co_Di is the most significant feature (0.7125), followed by C^1(0.439), H^3 (0.408), F^3 (0.372), and Y^1(0.250). For TV to SV class pair, Y^2 is the most significant feature (0.502) followed by corr_co_FP (0.4657), C^2(0.269), H^1 (0.364), and F^2 (0.337). For BS to Water class pair, F^2 is the most significant feature (1.943) followed by C^3 (1.499), Y^3 (1.392), corr_co_HD (0.8207), and H^1 (0.456). From the comparisons, it can be inferred that there is no clear indication of a set of features (from the same method) that are most significant for all class pairs. In short, there is no clear winner. Also, it is evident that the proposed PSCFs can be potential features to segregate land cover classes in mixed class scenarios.

Land cover classifications are performed using the PSCFs and other considered decomposition features in order to provide a comparative assessment based on LC classification performance. The classifier considered to be used with these features is support vector machines (SVM) with radial basis function (RBF) kernel because of its inherent optimum classification capability. All the classified images are shown in Figure 3.7. Overall classification performance indicators are provided in Table 3.4. It is evident from Table 3.4 that LC classification with PSCFs provides higher classification accuracy and kappa coefficient than the decomposition based features.

It is also evident from Table 3.4 that even if the proposed PSCF features performed better than the compared features, the classification accuracy can be further improved. For the proposed PSCF-SVM classification strategy, built-up (marker A) and TV (marker B) have low user accuracies (69.8% and 68.4% respectively) compared to other classes.

The classification performance can be further improved using more optimal class boundaries. Decision tree (DT) classifier provides the freedom to set optimal class boundary values during classification. Therefore, an optimal class boundary based DT classifier is considered and developed in this study. Details on the development of the optimal class boundary

estimation algorithm are discussed in section 3.5 of this chapter. Also, to reduce the computational complexity of the proposed classifier, feature selection is carried out with PSCFs.

It is clear from Table 3.2 that built-up can be segregated from all other classes (except tall veg) using corr_co_Di as a feature. Corr_co_Di is showing highest SI value amongst all the PSCFs. Further, it can be observed that (bare soil + water) are clearly separable from other classes (except short vegetation) using corr_co_FP as feature. With SI value of 0.83, bare soil and water can be separated using corr_co_HD. Water is also easily separable from other classes using corr_co_VD. Since both corr_co_HD and corr_co_VD can separate water, either or both of them can be used. It is evident from Table 3.2 that the cross-polarization PSCFs do not contain any complementary information for segregation process and hence are not considered for classification. A water canal (Ganga canal) is visible in Figure 3.1(b). Water in the canal is not identifiable in corr_co_HD or corr_co_VD (refer to Figure 3.8(a) and Figure 3.8(b)). A complementary feature is required which can address the issue. In Figure 3.1(a), the water canal is appearing blue (VV polarimetric band). Hence, VV can be used for classifying water in the canal. It is also evident from Figure 3.1(a), that only VV is not a sufficient choice for the classification of water as "bare soil" is not separable from water using only VV as feature. Hence, corr_co_HD or corr_co_VD, and VV polarimetric band are used. In summary, PSCFs selected for segregating a particular class based on significance are listed in Table 3.5.

Table 3.5 indicates that no PSCFs have been selected for segregating vegetation (both SV and TV) because of low SI values for the class pair over all the PSCFs. Observing Table 3.2, it is clear that the best choice among all the PSCFs for segregating tall and short vegetation is corr_co_FP. In Table 3.2, corr_co_FP is having the highest SI value among all PSCFs for both short and tall vegetation. Table 3.6 provides the updated and final list of selected PSCFs to be used during classification.

Table 3.2 Separability index (SI) values for various class pairs for all features.

Class pair	SI value							
	Corr_co_Di	Corr_cross_Di	Corr_co_FP	Corr_cross_FP	Corr_co_HD	Corr_cross_HD	Corr_co_VD	Corr_cross_VD
TV to BS	0.9427	0.0598	0.9940	0.3287	0.7855	0.4879	0.2188	0.4879
TV to Water	0.7928	0.0535	0.8634	0.3010	0.9089	0.4089	0.9546	0.4089
TV to Built-up	0.7125	0.1530	0.4930	0.2818	0.2953	0.5044	0.1452	0.5044
TV to SV	0.4333	0.0560	0.4657	0.0743	0.4249	0.1780	0.1244	0.1780
BS to Water	0.0596	0.0062	0.0280	0.0010	0.8207	0.0046	0.7306	0.3046
BS to Built-up	1.2884	0.2356	1.3986	0.529	0.5803	0.9359	0.3855	0.9359
BS to SV	0.4682	0.0113	0.4986	0.2269	0.3446	0.3765	0.0904	0.3765
Water to Built-up	1.1135	0.2264	1.2345	0.5015	0.9442	0.8352	0.9656	0.8352
Water to SV	0.3690	0.015	0.4230	0.2083	0.7516	0.2992	0.6806	0.2992
Built-up to SV	0.7968	0.2032	0.9014	0.3285	0.1806	0.6944	0.2787	0.6944

Table 3.3 Separability index (SI) values for various class pairs for popular features.

Class pair	SI value												
	C^1	C^2	C^3	Y^1	Y^2	Y^3	Y^4	F^1	F^2	F^3	H^1	H^2	H^3
TV to BS	0.744	0.780	0.500	0.833	1.635	0.313	0.851	0.040	1.733	0.357	0.763	0.460	0.804
TV to Water	1.080	1.139	1.828	2.260	1.763	0.706	0.833	0.847	2.625	1.670	0.232	0.173	0.328
TV to Built-up	0.439	0.001	0.428	0.250	0.009	0.216	0.230	0.209	0.118	0.372	0.241	0.146	0.408
TV to SV	0.008	0.269	0.276	0.050	0.502	0.011	0.171	0.013	0.337	0.067	0.364	0.010	0.022
BS to Water	0.368	0.480	1.499	1.364	0.543	1.392	0.148	1.350	1.943	1.466	0.456	0.349	0.333
BS to Built-up	1.174	0.670	0.079	0.905	1.481	0.577	1.002	0.306	1.744	0.649	0.469	0.477	1.056
BS to SV	0.688	0.533	0.090	0.831	1.244	0.319	0.704	0.069	1.481	0.364	0.396	0.347	0.716
Water to Built-up	1.496	0.993	1.105	1.925	2.351	0.422	0.970	0.563	2.562	1.647	0.005	0.181	0.603
Water to SV	0.939	0.882	1.202	1.928	0.801	0.731	0.663	0.953	2.408	1.445	0.108	0.018	0.273
Built-up to SV	0.393	0.235	0.117	0.173	0.453	0.227	0.377	0.214	0.538	0.245	0.105	0.152	0.361

C^1: Cloude-Pd (double bounce), C^2: Cloude-Pv (volume), C^3: Cloude-Ps (surface), (Cloude Decomposition)
Y^1: M-Y4D-Pd (double bounce), Y^2: M-Y4D-Pv (volume), Y^3: M-Y4D-Ps (surface), Y^4: M-Y4D-Pc (helix), (Modified Yamaguchi Decomposition)
F^1: GFD-Pd (double bounce), F^2: GFD-Pv (volume), F^3: GFD-Ps (surface), (Generalized Freeman Durden Decomposition)
H^1: H-A-Alpha-Entropy, H^2: H-A-Alpha-Anisotropy, H^3: H-A-Alpha-Alpha , (H-A-Alpha Decomposition)
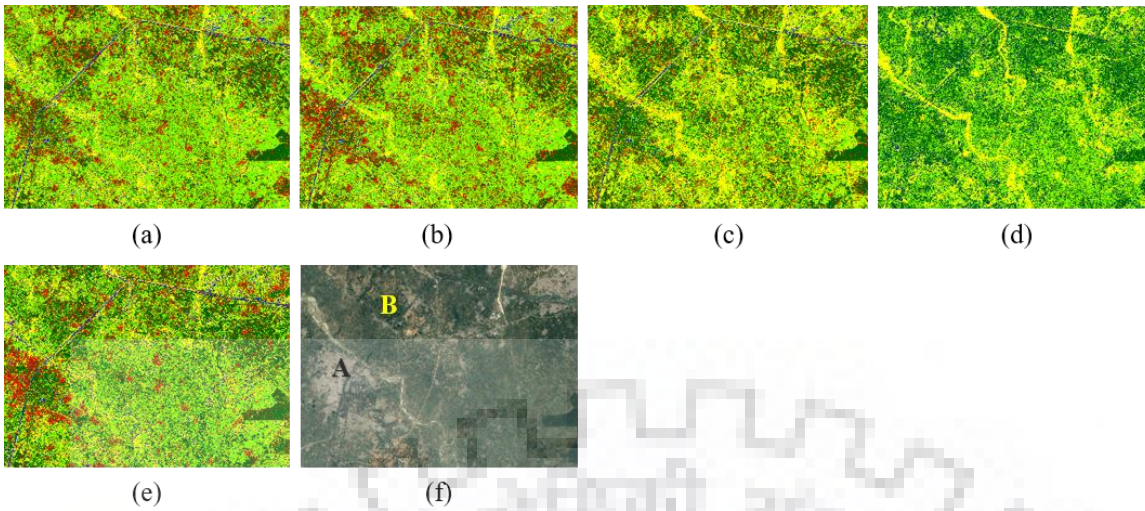
Figure 3.7: Classified Image; (a) M-Y4D-SVM, (b) GFD-SVM, (c) Cloude-SVM, (d) H-A-Alpha-SVM, (e) PSCF-SVM (proposed), and (f) Google Earth image for reference purpose. Colour scheme; red = built-up, light green = agriculture or SV, green = TV., blue = water, and yellow = bare soil. Marker A indicates built-up area and marker B indicates tall vegetation area.

Table 3.4 Classification performance of PS based and decomposition based PolSAR features. Classes: built-up (BU), water (W), bare soil (BS), short vegetation (SV), and tall vegetation (TV).

Method	Overall Accuracy (%)	Kappa	Class	User Accuracy (%)	Producer Accuracy (%)
H-A-Alpha-SVM	53	0.40	BS	63.2	82.9
			TV	44.6	88.9
			SV	58.8	53.3
			BU	14.3	1.3
			Water	47.7	12.8
M-Y4D-SVM	72.4	0.65	BS	66.1	82.3
			TV	70.9	88.0
			SV	76.9	77.5
			BU	61.5	50.3
			Water	88.9	54.8
GFD-SVM	71.9	0.63	BS	63.4	79.9
			TV	67.5	76.5
			SV	76.5	74.5
			BU	65.0	72.3
			Water	87.4	60.7
Cloude-SVM	63.7	0.53	BS	54.1	77.3
			TV	68.2	66.2
			SV	67.9	59.5
			BU	52.5	58.5
			Water	85.9	57.0
PSCF-SVM	74	0.69	BS	69.5	79.9
			TV	69.8	89.1
			SV	79	65.0
			BU	68.4	58.5
			Water	83.6	70.1

64
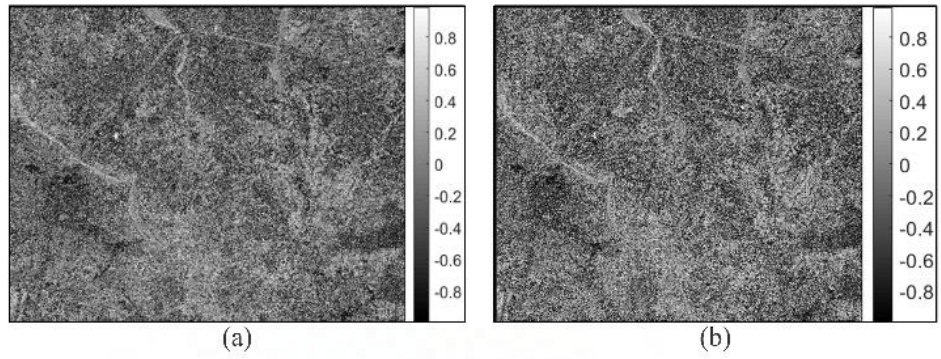
Figure 3.8 Correlation images (a) corr_co_HD and (b) corr_co_VD.

Table 3.5 List of selected feature corresponding to land cover classes.

Class	Feature selected for segregation
Built-up	Corr_co_Di
Water	Corr_co_HD, corr_co_VD
Canal	VV polarimetric band
Bare soil	Corr_co_FP

An alternate analysis of all the PSCFs as potential features for decision tree based land cover classification is presented in [179]. Scatter plot and OTSU multi- thresholding methods are used for separability assessment and similar results are reported. The class boundary values obtained using methods in [179] are rigid and do not change with the change in the PSCF statistics which in turn depends on the input PolSAR image. The changes captured in a PolSAR image can be via changes in landscape seasons or man-made. These changes needs to be considered while calculating class boundary thresholds. Therefore, in this study, a input image statistics based adaptive class boundary estimation approach is proposed to improve the results from [179]. Development of proposed approach is discussed in section 3.5.

3.5. Sub Task 2: Optimal Class Boundary Estimator based Decision Tree Approach for LC Classification

The approach is developed and presented in a sequential manner. Section 3.5 is broken down to four sub-sections and each sub-section is discussed in detail. Details of dataset are provided in section 3.3.2. Flowchart in Figure 3.9 summarizes the proposed algorithm.

3.5.1. PSCF statistics

Making the algorithm adaptive means the algorithm estimates class boundaries based on PSCF statistics (mean, standard deviation, etc.). Statistics for selected PSCFs i.e. mean (M) and standard deviation (S) are computed and are shown in Table 3.7. These statistics are used for the computation of class boundaries.

Table 3.6 Updated list of selected feature corresponding to land cover classes.

Class	Feature to be used for segregation
Built-up	Corr_co_Di
Water	Corr_co_HD, Corr_co_VD
Water Canal	VV polarimetric band
Bare soil	Corr_co_FP
Tall vegetation	Corr_co_FP
Short vegetation	Corr_co_FP

Table 3.7 Polarization signature correlation feature statistics.

Feature	Mean (M)	Std. dev. (S)
Corr_co_FP	0.5269	0.1266
Corr_co_Di	0.357	0.0716
Corr_co_HD	0.0580	0.1850
Corr_co_VD	-0.0580	0.1850
VV	-11.243	1.564

3.5.2. *Relationship between class boundaries and corresponding PSCF statistics*

A relationship is formulated between class boundaries and corresponding PSCF statistics (the one selected for segregating that particular class (refer Table 3.6). The relationship is formulated and explained with an example. Let's take the case of "built-up" class. Built-up is strongly segregated by "corr_co_Di" feature, $M_{corr\_co\_Di}$ and $S_{corr\_co\_Di}$ are mean and standard deviation of corr_co_Di. Therefore, built-up class spread can be expressed as;

$$M_{corr\_co\_Di} + n_1 * S_{corr\_co\_Di} \leq corr\_co\_Di \leq M_{corr\_co\_Di} + n_2 * S_{corr\_co\_Di} \qquad (3.5)$$

Where, n_1 and n_2 are real and scalar. $M_{corr\_co\_Di} + n_1 * S_{cor\_co\_Di}$ is the class lower boundary value and $M_{corr\_co\_Di} + n_2 * S_{corr\_co\_Di}$ is class upper boundary value. This expression means, if *corr_co_Di* pixel value lies in between these boundary values then the pixel belongs to built-up class. Similar expressions are developed for other classes (refer equation (3.6) through (3.9)).

- For Bare Soil

$$M_{corr\_co\_FP} + n_3 * S_{corr\_co\_FP} \leq corr\_co\_FP \leq M_{corr\_co\_FP} + n_4 * S_{corr\_co\_FP} \qquad (3.6)$$

- For Water

$$M_{corr\_co\_HD} + n_5 * S_{corr\_co\_HD} \leq corr\_co\_H \leq M_{corr\_co\_HD} + n_6 * S_{corr\_co\_HD} \qquad (3.7)$$

- For Short vegetation

$$M_{corr\_co\_FP} + n_7 * S_{corr\_co\_FP} \leq corr\_co\_FP \leq M_{corr\_co\_FP} + n_8 * S_{corr\_co\_FP} \qquad (3.8)$$

- For Tall vegetation

$$M_{\text{corr\_co\_FP}} + n_9 * S_{\text{corr\_co\_FP}} \leq \text{corr\_co\_FP} \leq M_{\text{corr\_co\_FP}} + n_{10} * S_{\text{corr\_co\_FP}} \tag{3.9}$$

Experiments suggest that class boundary thresholds for canal are not affected by VV band statistics and hence are not included for further study.

The goal of representing the class boundaries as in equation (3.5) through (3.9) is to obtain class boundaries that achieve best classification results under a given situation. Overall accuracy (OA) is an indicator of classification performance. Higher the OA, better the classification result. It is straightforward to understand that more accurate class boundaries result in higher OA. An empirical relationship is intuitive and can be developed between OA and class boundaries. The empirical relationship can then be treated as an optimization problem which can be addressed by an optimization technique. Since, M and S are constant for a given PSCF, therefore class boundary varies with only with n_i (i = 1 to 10). An empirical relationship between OA and n_i (i = 1 to 10) is relevant. Section 3.5.3 discusses the development of an empirical relationship between OA and n_i (i = 1 to 10).

3.5.3. Relating classification accuracy and class boundaries: Empirical modelling

In order to achieve the best overall accuracy (OA) with given feature space, the values of n_i (i = 1 to 10) should be optimized. To get optimal n_i (i = 1 to 10) values, an empirical function needs to be developed relating OA to n_i (i = 1 to 10) which can be optimized. OA is defined in equation (3.10).

$$OA = \frac{\text{total no of correctly classified pixels}}{\text{Total no of pixels tested}} \tag{3.10}$$

PSCFs are used based on Table 3.6 for segregating corresponding classes. Hence, for each class, individual overall accuracy (IOA) is computed and OA is rewritten as equation (3.11).

$$OA = \frac{IOA_{\text{built-up}} + IOA_{\text{bare soil}} + IOA_{\text{water}} + IOA_{\text{SV}} + IOA_{\text{TV}}}{5} \tag{3.11}$$

where, for example $IOA_{\text{tall\_veg}}$ is individual overall accuracy (IOA) in TV vs other class classification scenario using corr_co_FP as feature. Further, these IOAs can be empirically related to n_i (i = 1 to 10). Empirical relations developed between IOAs and n_i (i = 1 to 10) are

provided in Table 3.8. These relations are developed using surface fitting methods. For example, $\text{IOA}_{\text{tall\_veg}}$ is related to n_9 and n_{10} by a second degree polynomial surface.

$$\text{IOA}_{\text{TV}} = -573.7 - 998.9 * n_9 + 1742 * n_{10} - 1491 * n_9^2 - 27.5 * n_9 * n_{10}$$
$$- 1568 * n_{10}^2$$

(3.12)

Similar procedure is applied for all other classes and corresponding empirical relations are obtained. A coefficient of determination (R^2) value is also provided in Table 3.8. An optimization method can now be used to provide optimal n_i ($i = 1$ to 10) values that can maximize OA under constraints.

3.5.4. Optimizing the empirical model

Genetic algorithm (GA) [180] is used as an optimization tool for finding optimal n_i ($i = 1$ to 10) values. GA is a metaheuristic inspired by the process of natural selection that belongs to the larger class of evolutionary algorithms [181]. Genetic algorithms are commonly used to generate high-quality solutions to optimization and search problems by relying on biologically inspired operators such as mutation, crossover and selection. They imitate the biological processes of reproduction and natural selection to solve for the 'fittest' solutions. Like evolution, many of a genetic algorithm's processes are random, however this optimization technique allows one to set the level of randomization and the level of control. These algorithms are far more powerful and efficient than random search and exhaustive search algorithms yet require no extra information about the given problem.

Table 3.8 Empirical relations between individual overall accuracies (IOA) and n_i values.

Performance parameters	Derived empirical relation	R^2 (coefficient of determination)
$\text{IOA}_{\text{built-up}}$	$80.41 * \sin(5.3 * n_1 - 2.9) + 5.1 * \sin(33.5 * n_1 + 6.4)$	0.99
$\text{IOA}_{\text{bare\_soil}}$	$79.44 * \sin(5.5 * n_3 - 2.9) + 1.1 * \sin(69.2 * n_3 - 21.6)$	0.96
$\text{IOA}_{\text{water}}$	$2000 * \sin(4.7 * n_5 + 1.9) + 1993 * \sin(4.7 * n_5 - 1.3) + 1.23 * \sin(51.2 * n_5 + 2.2)$	0.92
IOA_{SV}	$-513.4 - 902.6 * n_7 + 2210 * n_8 + 93.5 * n_7^2 + 1250 * n_7 * n_8 - 1984 * n_8^2$	0.91
IOA_{TV}	$-573.7 - 998.9 * n_9 + 1742 * n_{10} - 1491 * n_9^2 - 27.5 * n_9 * n_{10} - 1568 * n_{10}^2$	0.99

The choice of GA as optimization method is because of the probability of more than one maxima/minima. Conventional methods like *first derivative* could be faster but can suffer from local maxima/minima problem.

The objective function of the optimization problem at hand is (refer equation (3.13);

$$\text{maximize} \quad (f(x)) = \text{maximize}(OA) \tag{3.13}$$

Subjected to

$$OA_{\text{initial}} \geq 50 \text{ \% (based on experiment)} \tag{3.13a}$$

$$-2.0 \leq n_1 \leq 2.0, n_2 = 2.0 \tag{3.13b}$$

$$-0.5 \leq n_3 \leq 1.0, n_4 = 1.0 \tag{3.13c}$$

$$n_5 = -2.0, -2.0 \leq n_6 \leq 1.0 \tag{3.13d}$$

$$-1.5 \leq n_7 \leq -1.0, -0.5 \leq n_8 \leq 0 \tag{3.13e}$$

$$-2.0 \leq n_9 \leq -0, -1.0 \leq n_{10} \leq -1.5 \tag{3.13f}$$

The upper and lower bounds of n_i (i = 1 to 10) values are obtained via experimental study during developing the empirical relation formulation as discussed in section 3.5.3 *and* are provided in equation 3.13(a-f). The internal parameters to GA are finalized using *trial and error* method.

After optimizing equation 3.13, the optimal n_i (i = 1 to 10) values obtained are shown in Table 3.9. These optimal n_i (i = 1 to 10) values are used to obtain class boundaries which in turn will be used in decision tree classifier in section 3.5.5.

3.5.5. *Optimal class boundaries based decision tree classifier*

The optimal n_i (i = 1 to 10) values obtained in section 3.5.4 are used to compute optimal class boundary values (using equation (3.5) through (3.9)) and henceforth, these class boundary values are used to develop a decision tree classification scheme. The decision tree based on the optimal class boundary values, in turn based on optimal n_i (i = 1 to 10) values is illustrated in Figure 3.10. The priority of a class to be segregated first is evaluated on how high is its SI value.

Table 3.9 Optimal n_i values obtained using the genetic algorithm.

n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}
0.315	2.0	0.446	1.0	-2.0	-1.667	-1.4	-0.2	-0.022	0.998
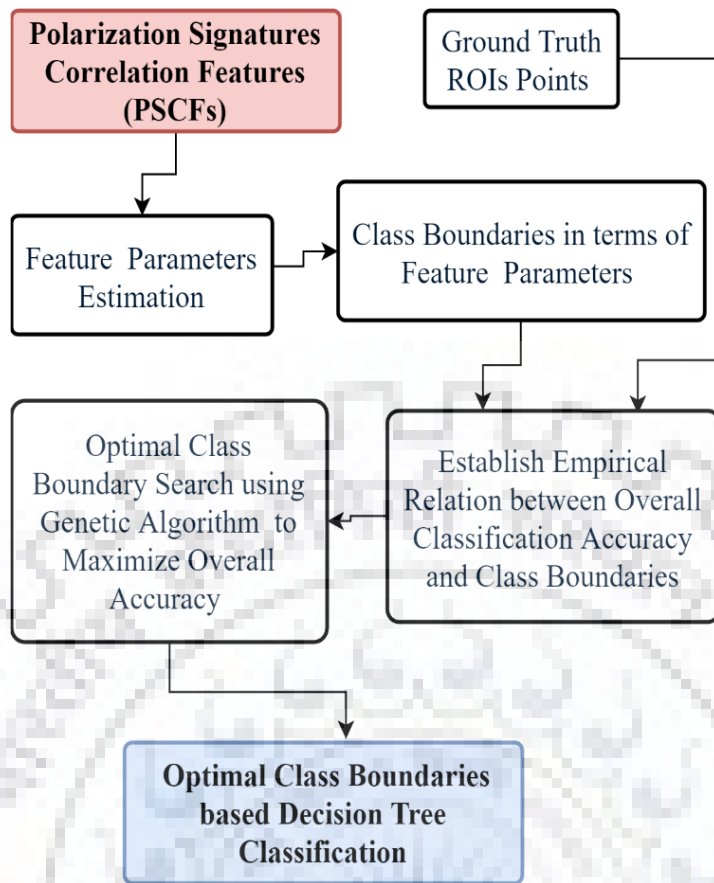
Figure 3.9 Flowchart illustrates the development process of the optimal class boundary estimator based decision tree classifier.
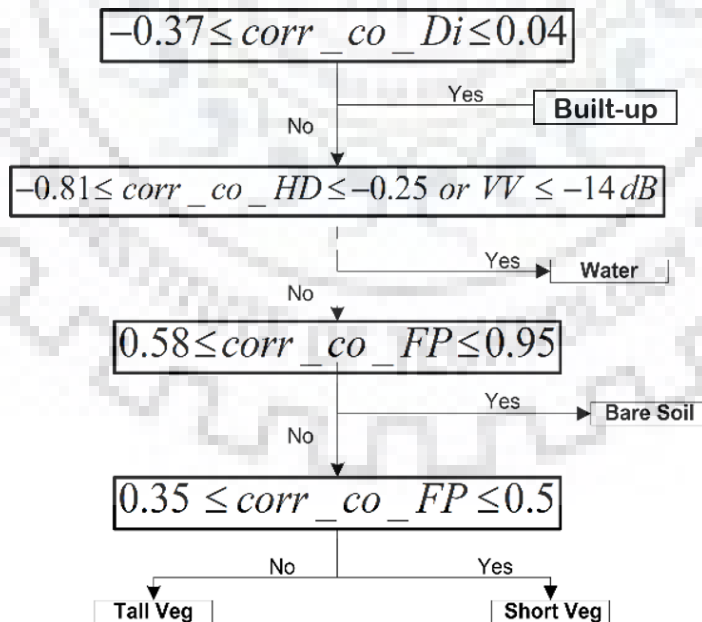
Figure 3.10 Decision tree based on optimal class boundaries values.

3.5.6. Results and discussion

The optimal class boundary estimator based decision tree classifier obtained in section 3.5.5 (refer Figure 3.10) is utilized to perform LC classification on Roorkee dataset. The classified image is shown in Figure 3.11(a). The proposed optimal class boundary estimator based DT classifier is compared with the RBF-SVM classifiers. The same PSCFs are used as features with the RBF-SVM. Classified image the RBF-SVM is shown in Figure 3.11(b) respectively. Overall classification performance indicators for the proposed method and RBF-SVM are provided in Table 3.10. From Table 3.10, it is evident that the proposed method achieves best classification accuracy (75.3%) with kappa coefficient (0.69). Also, the user accuracy for built-up and TV have improved to 73.8% and 72.4% respectively (Marker A, and B respectively). Markers A and B in Figure 3.1 are clearly separated using the proposed optimal class boundary approach (refer figure 3.11(a)).

However, computing PSCFs from PSs for further utilization with classifiers seems an extra step. It is motivating to know that two dimensional CNNs (2D-CNNs) are powerful image classifiers which can extract complex features directly from data. Since, the PSs are two dimensional representation of radar backscatter response of a land cover (refer section 3.2.1), they can be used as input images for pixel wise LC classification. The idea of utilizing CNNs with PSs for LC classification is proposed in section 3.6.

Table 3.10: Classification performance indicators for the proposed method and RBF-SVM classifier with PSCFs as features.

Method	OA (%)	Kappa	Class	User Accuracy (%)	Producer Accuracy (%)
Proposed optimal class boundary estimator based DT classifier	75.3	0.69	BS	69.5	79.9
			TV	73.8	89.1
			SV	78.9	65.0
			BU	72.4	71.0
			Water	83.6	70.0
RBF-SVM	74	0.69	BS	69.5	79.9
			TV	69.8	89.1
			SV	79	65.0
			BU	68.4	58.5
			Water	83.6	70.1

Figure 3.11: Classified image: (a) Optimal class boundary estimator based DT classifier, (b) RBF-SVM, and (c) Google Earth image for reference purpose. Colour scheme; red = built-up, light green = SV, green = TV, blue= water, and yellow = BS. Marker A indicates built-up area and marker B indicates tall vegetation area.

3.6. Sub Task 3: Direct Utilization of PolSAR Polarization Signatures via Convolutional Neural Networks for LC Classification

The current section presents utilization and analysis of advanced pattern classification methods for PolSAR land cover classification. Convolutional neural networks (CNNs) are recently the most popular, efficient, and powerful techniques in image feature extraction and classification applications. Their parameter sharing and local connectivity properties makes them best tools for extracting information from images [35]. CNNs in combination with deep learning technology are being widely used in PolSAR data applications [58]. In this study, CNNs are utilized to extract features from PSs and perform land cover classification. A CNN based PSs pattern classification technique is proposed here which considers the influence of the neighbourhood of a pixel and thereby utilizing the PSs directly.

3.6.1. Two dimensional convolution over polarization signatures: The concept

Figure 3.12 demonstrates the utilization of two dimensional convolutions on PolSAR PSs. Figure 3.12(a) and Figure 3.12(b) depicts isometric view and top view respectively of the PS of a PolSAR image pixel. A 2D convolution filter as shown in Figure 3.12(d) convolves over the PS in a sliding window manner and provides convolutional features. These features can be used in classification, regression or recognition.

72
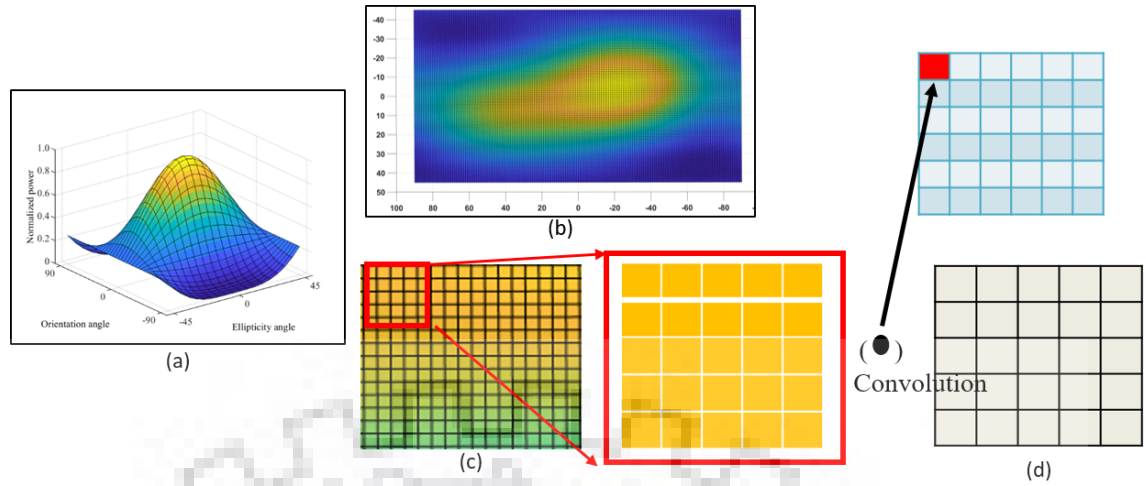
Figure 3.12: Two dimensional convolution over polarization signature. (a) Polarization Signature of a PolSAR image pixel. (b) Top view. (c) Polarization signature subset. (d) Convolutional filter.

3.6.2. PSs-CNN model architecture design and training

A CNN model is constructed and configured with the help of convolutional and fully connected layers to perform land cover classification with PSs. The proposed CNN model is termed as PSs-CNN model. The proposed PSs-CNN model configuration is provided in Table 3.11. The input to the PSs-CNN model is a pixel PSs. Pixel (observed) PSs used here are obtained as discussed in section 3.4.1. Each PS corresponds to a 19×37 image. The 19×37 input image has two channels i.e. *copol* and *crosspol*. Therefore, overall the size of input to the PSs-CNN is $19 \times 37 \times 2$. At first, the PSs-CNN used four convolutional layers. Number of convolution filters and filter sizes used in each layer are listed Table 3.11. ReLU activation function and dropout fraction of 10% is used in each convolutional layer. The dropout regularization strategy is employed to achieve less overfitting and better generalization. Further, two fully connected or FC layers, first with 50 nodes and second with 10 nodes are considered. Both FC layers used the ReLU activation function and a dropout fraction of 10%. Finally, a softmax classification layer with 5 labels (based on 5 output land cover classes) is used. A batch normalization process is carried out before the convolutions to speed up the training process and to reduce the network sensitivity. A seven layered PSs-CNN classification model is realized as shown in Figure 3.13.

The proposed PSs-CNN model is trained on labelled samples provided in Table 3.1. The loss function considered during training is the categorical cross entropy function. The learning rate is set to 0.01 and the number of training epochs is set to 200. Adam optimizer is used for network weight optimization.

Table 3.11 Summary of PSs-CNN model configuration, hyperparameters and training

Layer →	Input	Convolutional layers; C1, C2, C3, C4,C5	Fully connected layers, FC1, FC2	Output layer
PSs- CNN Architecture	Input image size = $19 \times 37 \times 2$ Normalization = Batch Normalization	Filters = 10, 10, 10, and 20 Kernel size = 3×5, 3×5, 3×5, and 9×21 Activation = ReLU, ReLU, ReLU, and ReLU Dropout fraction = 0.1, 0.1, 0.1, and 0.1 Pooling size = 0, 0, 0, 0, 0,	Nodes =50,10 Activation = ReLU, ReLU Dropout fraction = 0.1, 0.1	Labels = 5 Activation = Softmax
PSs- CNN Training parameters	Loss function = Categorical cross entropy Learning rate = 0.01 Number of epochs = 200 Optimizer = Adam			
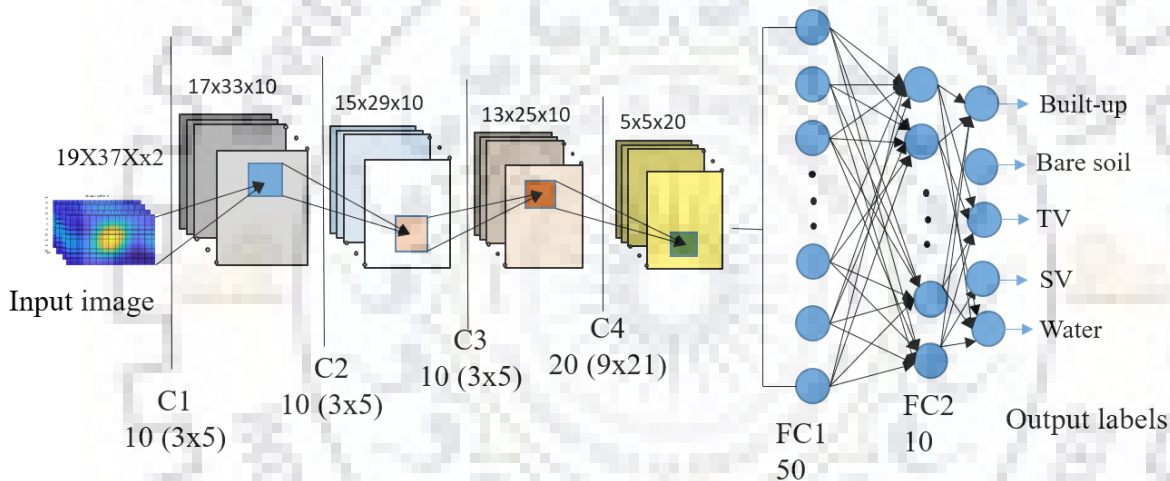
Figure 3.13 Architecture of the proposed PSs-CNN model.

3.6.3. Results and discussion

Figure 3.14(a) is the classified image for using the PSs-CNN model as classifier. To assess the classification performance of proposed PSs-CNN model, the PSs-CNN model performance is compared with two popular PSs based land cover classifiers i.e. the normalized signature correlation mapper (NSCM) based minimum distance classifier and the normalized Euclidean distance (NED) based minimum distance classifier. Figure 3.14(b) is classified image obtained using the NSCM classifier whereas Figure 3.14(c) is the classified image obtained using the NED classifier. It is evident form Figure 3.14 that PSs-CNN outperforms the NSCM and NED based classifiers. Table 3.12 summarizes the quantitative assessment of classification performance of PSs-CNN, NED, and NSCM classifiers. The PSs-CNN classifier performs land cover classification with good overall accuracies, i.e. 72.8% with kappa coefficient of 0.69. In contrast,

the NED performed the classification with 62.3 % and the NSCM performed the classification with 64.1%. Both NED and NSCM showed low kappa values (refer Table 3.12). Because of the point-wise operability, the NED and NSCM classifiers are not able to capture and exploit the neighbourhood relationship of a pixel and thereby are not able to efficiently address the classical classification challenges with PolSAR data such as, Built-up vs Tall veg, where volume scattering is dominant [182]. The proposed PSs-CNN classifier resolves the issue significantly by extracting and utilizing the subtle differences in the PSs of built-up and forest. Markers A and B (refer Figure 3.1(a)) are clearly separated using the proposed approach (refer Figure 3.14(a)).

Table 3.12 Classification performance indicators for the PSs-CNN, NSCM, and NED classifiers with PSs as features. BS = bare soil, TV = Tall vegetation, SV = Short vegetation, BU = Built-up.

Method	OA (%)	Kappa	Class	User Accuracy (%)	Producer Accuracy (%)
PSs-CNN	72.8	0.69	BS	88.5	79.9
			TV	73.8	89.1
			SV	62.9	65.0
			BU	70	58.5
			Water	79.6	70.1
NSCM	64.1	0.54	BS	100.0	61.5
			TV	100.0	44.5
			SV	65.6	82.2
			BU	38.5	57.6
			Water	56.2	73.2
NED	62.3	0.51	BS	88..4	61.1
			TV	100.0	44.0
			SV	65.5	82.2
			BU	33.8	47.0
			Water	56.2	73.2

(a)　　　　(b)　　　　(c)

(d)
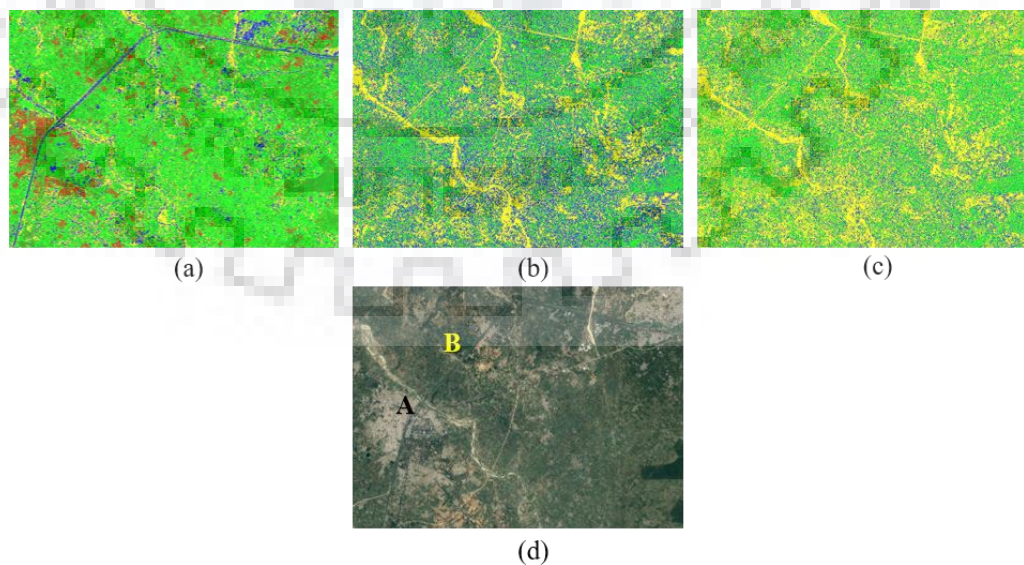
Figure 3.14 Classified Image: (a) PSs-CNN, (b) NSCM, (c) NED, and (d) Google earth image. Colour scheme; red = built-up, light green = SV, green = TV, blue= water, and yellow = BS. Marker A indicates built-up area and marker B indicates tall vegetation area.

3.7. Overall Results and Discussion

Comparison between the two novel approaches developed in this study i.e. the optimal class boundary estimator based DT classifier and the PSs-CNN is made. Classification results from the two novel classifiers and the RBF-SVM are summarized in Table 3.13. The comparison indicated that the optimal class boundary estimator based DT classifier is showing an approx. 3% improved classification accuracy than PSs-CNN classifier. However, overall comparison suggest that PSs-CNN is easier to implement, does not require an extra step (computing PSCFs), and is scalable.

Table 3.13 Classification performance indicators for the optimal class boundary estimator based DT, PSs-CNN, and RBF-SVM classifier.

Method	OA (%)	Kappa
PSs-CNN with polarization signatures	72.8	0.69
Optimal class boundary estimator-DT with PSCFs	75.3	0.71
RBF-SVM with PSCFs	74	0.70

3.8. Conclusions

Overall a study where an attempt is made to extract information from one of the most popular satellite image modality i.e. PolSAR images is presented in this chapter. The study is composed of three sub-tasks.

Under the first sub task i.e. *sub task 1*, the investigation of polarization signatures (PSs) as potential PolSAR features for land cover (LC) classification in mixed class scenarios is done. Novel PSs based features i.e. PSCFs are computed based on the degree of similarity between observed and canonical PSs. Further, PSCFs are analysed, and are compared with decomposition based features for the purpose of LC classification. Comparisons indicate potential of PSCFs as significant features for LC classification in mixed land cover class scenarios.

Second sub task i.e. *sub task 2*, covers successful development and implementation of an optimal class boundary estimator based DT classifier which exploits the PSCFs to achieve improved LC classification. During the development of the optimal class boundary estimator based DT classifier, image and individual class-wise PSCFs statistics are utilized to formulate empirical relationship between classification performance indicator (overall accuracy) and class boundary values. Optimization of this empirical relationship with genetic algorithm resulted in optimal class boundaries which are utilized in a DT to perform LC classification. Classification results using this approach are compared with classification results from RBF-SVM, while keeping the features identical i.e. PSCFs. Classification performance of the proposed approach

especially user and producer accuracies for built-up class (72.4% and 71.0% respectively) and tall veg class (73.8%, and 89.1% respectively) indicate that the proposed approach, i.e. the optimal class boundary estimator based DT classifier, achieved better classification results in mixed LC class scenarios.

In the third sub task i.e. *sub task 3*, development of a two dimensional CNN (2D-CNN) based novel model for LC classification is presented. The 2D-CNNs can directly exploit PSs to perform LC classification, and therefore, the approach reduces the efforts to compute PSCFs. Also, since 2D-CNNs are powerful image extractors, they are able to extract the subtle contextual patterns present in PSs of different LCs. In summary, a novel PSs-CNN LC classification model is realized. Classification results from this PSs-CNN model are compared with NED and NSCM based minimum distance classifiers which are popularly used with PSs for land cover classification. Comparison indicated a superior performance of the novel PSs-CNN model for land cover classification.

Comparison between the two novel approaches proposed in this study i.e. the optimal class boundary estimator based DT classifier and the PSs-CNN model based classifier is made. The comparison indicated that the optimal class boundary estimator based DT classifier is showing an approx. 3% improved classification accuracy than PSs-CNN classifier. However, the PSs-CNN model is easier to implement, does not require an extra step (computing PSCFs), universally applicable, and is scalable.

Therefore, the study presented in this chapter concludes that; i) PSs have shown significant potential as PolSAR features for land cover applications, ii) segregation of built-up class and tall vegetation class in PolSAR data can be improved by using optimal class boundary estimators with decision trees for classification, and iii) 2D-CNNs have shown potential as classifiers with PSs for LC classification in terms of ease of implementation, and scalability. However, the potential of PSs-CNN model needs to be further investigated for other important PolSAR data based earth observation applications.

Chapter 4

NOVEL APPROACH FOR DETECTION OF HUMAN SETTLEMENTS AFFECTED DURING FLOODS USING POLARIZATION SIGNATURES AND CNNS

During floods, an updated and accurate information on flood affected human settlements helps mitigate the disaster efficiently. This chapter presents an efficient approach to map human settlements, especially villages in the rural parts of India that are affected during flood using PolSAR imagery, convolutional neural networks (CNNs), and image morphology. This chapter is an extension of chapter 3 in a sense to utilize the PSs-CNN model (proposed in chapter 3) for a crucial, and practical application.

The approach involves three stages: –*first*, built-up area extraction from a PolSAR image, –*second*, detection of villages in a built-up area image, and –*third*, identification and mapping villages that are affected by flood. In the first stage, a CNN based novel classification model is proposed to extract built-up areas from PolSAR image. This model utilizes PolSAR polarization signatures (PSs) as features. The second stage proposes a novel village detection methodology. This methodology detects whether a particular object in the built-up area image (obtained in the first stage) is a village or not. In the third stage, flood affected villages are identified and mapped and non-flooded villages are masked out using another image morphology based methodology.

4.1. Introduction

Floods are the most frequent, widespread, and impactful disasters in India. A major population of India lives in towns and villages and during the floods, rescue operations to these affected areas are both difficult and challenging. Rapid and reliable management and mitigation is key to minimize the loss of human lives and property during a flood disaster. Accurate and updated information of flood affected area can lead to an efficient disaster mitigation plan. In the context, satellite images based flood disaster monitoring is one of the most helpful and reliable technology to counter these flood disasters. Especially, synthetic aperture radar (SAR) imaging has proved to be a very successful and promising technology for both, monitoring floods or, planning rescue operations in flood affected areas [16], [183]–[186]. Polarimetric synthetic aperture radar (PolSAR) imaging sensors are developed to capture the changes in polarization states introduced by the interaction of EM waves with earth surface features (land covers). Among many other

applications, PolSAR imaging has also proved its significance and uniqueness in flood monitoring and mitigation [183], [187]–[189]. PolSAR data is usually captured in four *linear* polarization states namely HH, HV, VH, and VV. The scattering responses in different polarization states help the identification and classification of different land covers which is helpful during flooded area assessment. For example, built-up structures usually show double bounce scattering response. This means that these built-up structures appear bright (high scattering response) in HH PolSAR image. On the other hand, tall vegetation areas (e.g. forests) show multiple bounce scattering (volume scattering) response, or smooth surfaces such as water show single bounce scattering response [190]. However, the PolSAR data has been transformed, represented, and utilized in different other forms and bases in various flood monitoring studies [191], [192]. Each PolSAR data representation has its advantages and disadvantages and its selection depends on application and available processing methodologies.

Monitoring human settlements in the rural parts of India during floods is difficult even with PolSAR data. The built-up structures (e.g. houses) in these human settlements, especially villages of rural India are closely packed and randomly oriented. Therefore, these villages show volume scattering response along with double bounce scattering response in a PolSAR image. The mixed scattering response from these villages hinders their segregation from tall vegetation in a PolSAR image. This challenge is also discussed in chapter 3 (refer section 3.1). Therefore, the detection and mapping of these villages during a flood using PolSAR image is challenging as there is high risk of false alarms, or missing out a flood affected village during rescue operations. The current study considers PolSAR polarization signatures (PSs) as features to improve the detection and mapping of villages that are affected by flood. Polarization signature (PS) is a two dimensional representation relating backscatter response from a target at a particular polarization state. Polarization signatures or PSs are introduced in [155] and later on represented graphically by many researchers in different ways [154]. PSs have been proved to be useful for characterization of land covers such as built-up, forest, or water [149]. A comparison of PSs based features and decomposition based features for land cover characterization is also presented in section 3.4. The study demonstrates the significance of the utilization of PSs for land cover characterization and classification in mixed land cover scenarios which are similar to the scenario at hand in the current study. Therefore, the utilization of PSs as PolSAR features is considered in the current study. A detailed description on PSs is provided in section 3.2.1.

Moreover, it is also evident from most PSs based land cover studies that these studies utilized distance/correlation measures between canonical target PSs and observed PS for detection, identification, and classification of land covers [149], [153], [193] Normalized signature correlation mapper (NSCM) [194], and normalized Euclidean distance (NED) [193]

are the popular distance/correlation based classifiers used with PSs for various applications. There is a lack in PSs based studies because of the fact that PSs are two dimensional features whereas the common processing techniques used with these PSs such as the NSCM or the NED, perform vector operations, and therefore are unable to consider the neighbourhood influence during target (land cover) identification or classification. Hence, the PSs are under-utilized and provide sub optimal results.

The current study also addresses this challenge of under-utilization of PSs and proposes a novel approach based on convolutional deep neural networks to fully exploit the PSs as land cover features. Convolutional deep neural networks (CNNs) are recently one of the most popular, efficient, and powerful techniques in image feature extraction and classification applications. Their parameter sharing and local connectivity properties makes them best tools to extract complex and localized features from images [35]. A brief introduction to CNNs is provided in chapter 3. In the recent years, CNNs in combination with deep learning technology are being widely used in PolSAR data applications [58]. The utilization of CNNs with PSs for land cover classification has been introduced in chapter 3. The study in chapter 3 indicates the significance of CNNs as PSs feature extractors. The study also reports improvement in classification performance with utilization of CNNs with PSs for land cover classification. Therefore, the study presented here in chapter 4 explores and extends the utilization of CNNs with PSs in order to extract built-up areas from a PolSAR image.

The study also utilizes *apriori* knowledge and image morphology to build village detection filters and to mask out misclassifications. Image morphology has been extensively used in studies to detect built-up areas from satellite images [195].

To summarize, the study in this chapter presents a novel approach to detect and map villages affected by a flood with the help of PSs, CNNs, *apriori* knowledge and image morphology. The approach presented here involves three stages. These are;

—*first*, design and development of PSs based CNN model for extraction of built-up areas from PolSAR image which contains the flooded region,

—*second*, development of novel methodology for design of village detection filters based on *apriori* knowledge and image morphology, and

—*third*, development of novel methodology for identification, and mapping of flood affected villages using image morphology.

This chapter is divided into following sections. Section 4.2 provides detailed information on the experiment setup, study areas and datasets considered and, ancillary information collected for the study. Section 4.3 discusses the methodologies developed in the three stages of the study.

Qualitative and quantitative evaluation of the results from each methodology is also done and discussed in this section. At last, section 4.4 concludes the study.

4.2. Experiment Setup

4.2.1. Study area

Two areas are considered for the analysis, development and testing of the proposed methodology. The first study area is common from chapter 3. This study area is termed as 'Roorkee' study area in this chapter. Details of this study area are already provided in section 3.3.1. The second study area includes Haridwar city of the northern India and neighbouring region. The area extends from $29°58'23''N$, $77°49'52''E$ to $29°49'23''N$, $78°10'5''E$ The area covers about 270 km² of landscape. This second study area is termed as 'Haridwar' study area in this chapter. Both study areas have mixed class type landscape, which is beneficial for the evaluation of classifiers performance in mixed class scenarios. Figure 4.1(a) and Figure 4.2(a) shows a Google Earth image of the Roorkee and Haridwar study areas respectively.

4.2.2. Satellite data

PolSAR data used in chapter 3 (refer section 3.3.2) is considered here as well for experiment purpose. Preprocessing is required with raw PolSAR data as it suffers from artefacts such as speckle noise. Preprocessing of the PALSAR-2 PolSAR data is similar to as explained in chapter 3 (refer section 3.3.4). Four features namely HH, HV, VH, and VV are obtained. Since, HV and VH are similar due to the reciprocity theorem, only HH, HV, and VV are used further. Figure 4.1(b) shows a false colour image of the Roorkee dataset and Figure 4.2(b) shows a false colour image of the Haridwar dataset

4.2.3. Ground truth

Ground truth is collected on two counts. First, ground truth sample points are collected for classifier training and validation. Five land cover classes are considered namely bare soil (BS), tall vegetation (TV), short vegetation (SV), built-up, and water. Ground truth class sample points are measured directly on the terrain based on visual inspection. Approximately 1200 data points are collected for both the study areas. Details of the ground truth collected for both study areas are shown in Table 4.1.

Second, the total number of villages (a common term opted here to represent all types of human settlements in the study region) in each study area are counted for validation of the

proposed approach. Table 4.1 also lists the actual number of villages in the Roorkee and Haridwar study areas. Further, Figure 4.3(a, and b respectively) highlights (in white colour) the villages in the Roorkee and Haridwar study regions captured from Google Earth imagery via a digitization process. Any village in the rural parts of India consists of a minimum of 20 houses and is spread over a minimum area of 7600-meter square. These physical dimensions help in detection of villages in a PolSAR image.

It is important to note here that in this chapter; different terminologies have been used at different stages to represent different versions of built-up areas. Basic terminologies have been defined here.

- Built-up structure: Man-made structures.
- Built-up area: A collection of built-up structures.
- Village or town: Built-up area larger than a predefined size.
- Object: A connected group of image pixels.
- Detected villages: Villages that have been successfully detected.
- Flooded villages: Villages that are affected by flood.

These terminologies help to better understand the different stages of the study.

Also, in order to evaluate the performance of proposed approach, artificial flood scenarios are considered and used during the study. Details of the process of creating artificial flood scenarios are provided in section 4.3.3.2.
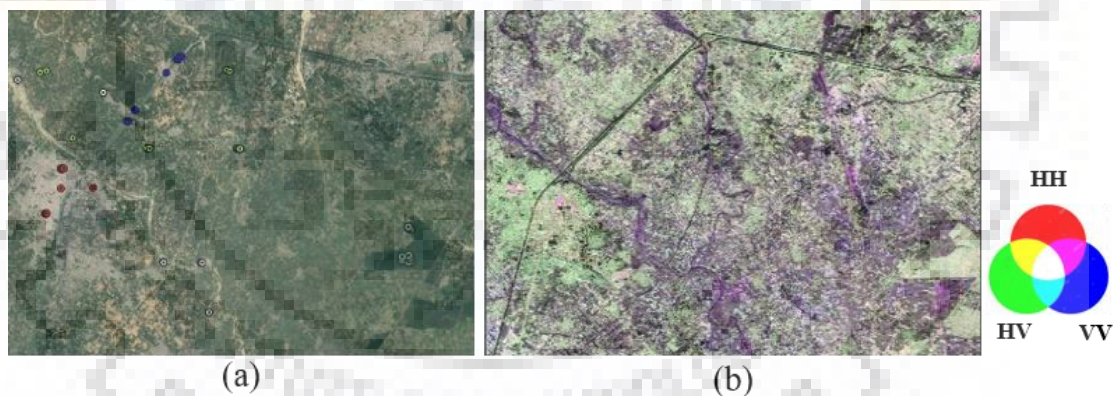
Figure 4.1 (a) Google Earth imagery of the Roorkee with markers indicating major land cover classes. Markers highlighted in dots: red = built-up, green = TV, blue = water, light green = SV, and yellow = BS (Courtesy: Google Earth).   (b) False colour representation of preprocessed SAR data.
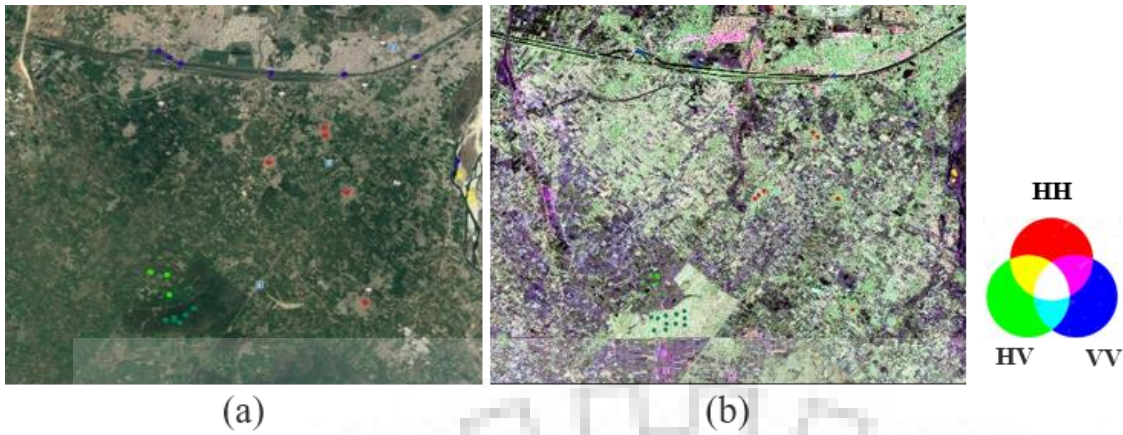
Figure 4.2 (a) Google Earth imagery of the Haridwar with markers indicating major land cover classes. Markers highlighted in dots: red = built-up, green = TV, blue = water, light green = SV and yellow = BS (Courtesy: Google Earth). (b) False colour representation of preprocessed SAR data

Table 4.1. Summary of ground truth samples collected for the study. *BS = bare soil, TV = dense tall vegetation, SV = short vegetation. All points are acquired directly on the field by visual inspection at selected GPS coordinates. Also, the total number of villages in both the study regions are listed here.

Land Cover	Roorkee		Haridwar	
	Training and validation	Testing	Training and validation	Testing
BS	200	40	200	40
TV	200	40	200	40
SV	200	40	200	40
Built-up	200	40	200	40
Water	200	40	200	40
Total	1000	200	1000	200
Total number of villages in Roorkee region		93		
Total number of villages in Haridwar region		60		
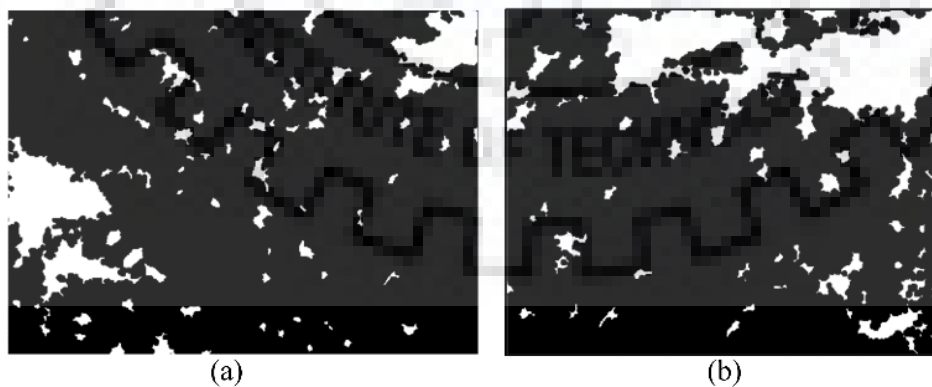
Figure 4.3 Areas highlighted in white are villages in the (a) Roorkee, and (b) Haridwar study areas. Village areas are extracted from Google Earth imagery of the corresponding regions. (Courtesy: Google Earth).

84

4.3. Proposed Approach, Implementation, and Discussion

This section discusses the development of a novel approach to detect, identify, and map flood affected villages in a PolSAR image. The overall approach is divided into three stages.

First, design and development of a novel PSs based CNN model for extraction of built-up areas in PolSAR image. The design and development of the novel PSs based CNN model is discussed in sections 4.3.1 and 4.3.2. Model performance evaluation and comparisons are discussed in section 4.3.2.

Second, development of a novel methodology for design of village detection filters using *apriori* knowledge and image morphology. These filters help in detection of villages in a built-up area image and also help suppress false alarms or misclassifications. Section 4.3.3.1 covers the development of this methodology. Qualitative and quantitative evaluation of the detection performance of this methodology is also covered in this section.

Third, development of methodology for identification and mapping of villages that are affected by flood and masking out villages that are not affected by floods. A novel methodology, which exploits the phenomena of enclosing of a village by water if it's flooded, is developed. Section 4.4.3.3.2 discusses the development of this methodology. Evaluation of this novel methodology is discussed in section 4.4.3.3.2. The three methodologies proposed in the three stages sum up to provide a novel flood affected village detection and mapping approach. Flowchart shown in Figure 4.4 depicts the overall proposed approach.

4.3.1. Land cover polarization signatures

According to the discussion in section 3.2.1, both, co-polarized and cross-polarized PSs are computed from pre-processed PolSAR data using equation (3.1). Co-polarized PS plots for considered land cover class (averaged over all class samples, refer to Table 4.1) are shown in Figure 4.5. Each PS plot in figure 4.5 shows backscatter power response for ellipticity angle χ ranging from -45° to $+45^o$ and orientation angle ψ ranging from -90° to $+90^o$ with a step size of 5° in each direction resulting in a 18×37 two dimensional matrix. It is evident from Figure 4.5 that PSs of considered land covers are different from each other.

These PSs are used as features for land cover characterization and classification in many studies. Classification schemes used in [149], [179], and [193] for PolSAR PSs based land cover classification are distance/correlation based which do not consider the neighbourhood relationship of a pixel during classification and therefore PSs remain under-utilized. There is a need to explore and develop advanced classifiers that can capture local spatial relationship of a PS pixel during classification which might improve the classification. A CNN based PSs

classification technique is introduced in chapter 3 which considers the influence of the neighbourhood of a PS pixel and thereby utilizes the PSs efficiently. The current study extends the utilization of PSs with CNNs to extract built-up areas from PolSAR image.
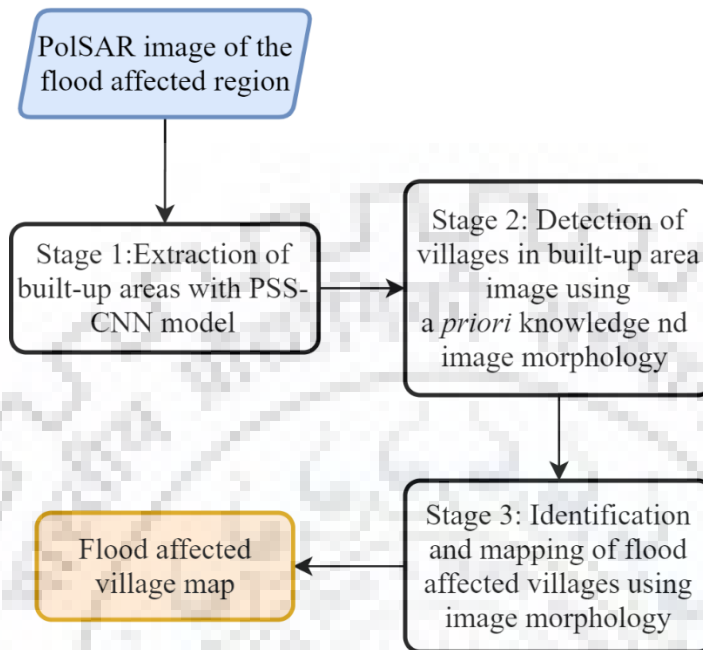
Figure 4.4 Flowchart of the proposed approach
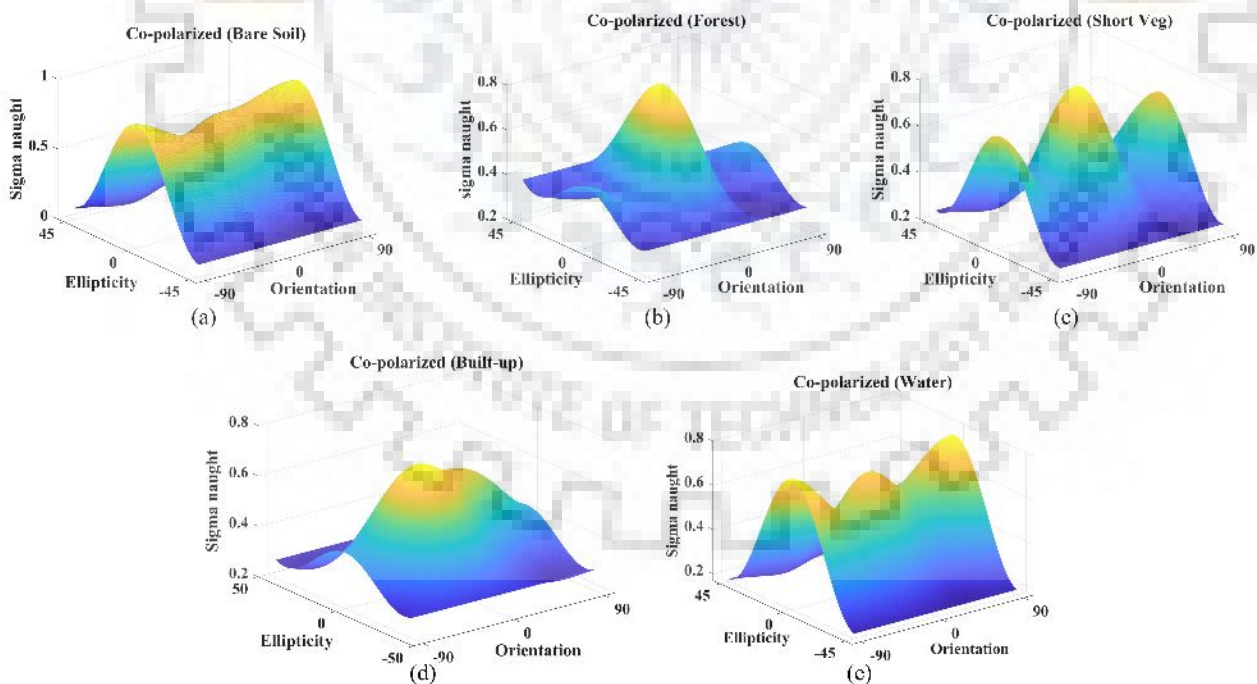
Figure 4.5 Sample co-polarized PSs of land cover classes mentioned in Table 4.1. (a) bare soil, (b) TV, (c) SV, (d) built-up, and (e) water.

4.3.2. Stage 1: CNN and polarization signatures based classification model for built-up area extraction

In this stage, CNNs are utilized to extract features from PSs for built-up area extraction. A CNN based PSs classification technique is first introduced in chapter 3. The technique was able to capture the influence of the neighbourhood of a PS pixel and thereby utilizing the PSs more efficiently. A novel CNN model similar to the model discussed in section 3.6 is developed here with the help of convolutional and fully connected, layers. The proposed CNN model is termed as PSs-CNN model from here onwards in this chapter. The proposed PSs-CNN model configuration is provided in Table 4.2. Unlike the CNN model in chapter 3 (refer section 3.6) where per pixel PS (polarization signature of one pixel) is used for classification, a window of PSs is used as input here. The window based approach helps suppresses the random volume scattering responses from closely packed built-up structures in built-up areas and thereby improves the segregation of these built-up areas. The variation in PS statistics with varied window sizes are listed in Table 4.3. From Table 4.3, it is evident that as the window size increases, the range of the peak value for each class decreases. This indicates to the suppression of scattering response anomalies around the considered pure class pixel. This in turn improves the training and classification performance of the PSs-CNN model.

A 3×3 window of neighbouring PSs is used here as input. This corresponds to a 54×111 (refer section 4.3.1) image as input to the PSs-CNN model. This 54×111 input image has two channels i.e. copol and crosspol. Therefore, the overall size of input to the PSs-CNN is $54\times111\times2$. The PSs-CNN used five convolutional layers. The number of convolution filters and filter sizes used in each layer are listed in Table 4.2. ReLU activation function and dropout fraction of 10% is used in each convolutional layer. The dropout regularization strategy is employed to reduce overfitting and achieve better generalization. Further, two Fully Connected or FC layers, first with 50 nodes and second with 10 nodes are used. Both FC layers used the ReLU activation function and a dropout fraction of 10%. Finally, a Softmax classification layer with 3 output labels is used. The first output class is the built-up class which needs to extracted, the second is 'rest' class (flood), and the third output class is TV class because TV and built-up are the only classes that are not completely submerged in water during a flood. A batch normalization process is carried out before the convolutions to speed up the training process and to reduce the network sensitivity. Overall, an eight layered PSs-CNN classification model is realized as shown in Figure 4.6.

Table 4.2 Summary of PSs-CNN model configuration, hyperparameters and training

Layer →	Input	Convolutional layers C1, C2, C3, C4, C5	Fully connected layers FC1, FC2	Output layer
PSs- CNN Architecture	Input image size = 57*111*2 Normalization = Batch Normalization	Filters = 10, 10, 10, 15, 15, Kernel size =9×13, 9×13, 9×13, 15×33, 15×33 Activation = ReLU, ReLU, ReLU, ReLU, ReLU Dropout fraction = 0.1, 0.1, 0.1, 0.1, 0.1 Pooling size = 0, 0, 0, 0, 0,	Nodes =50,10 Activation = ReLU, ReLU Dropout fraction = 0.1, 0.1	Labels = 3 Activation = Softmax
PSs- CNN Training parameters	Loss function = Categorical cross entropy Learning rate = 0.01 Number of epochs = 200 Optimizer = Adam			
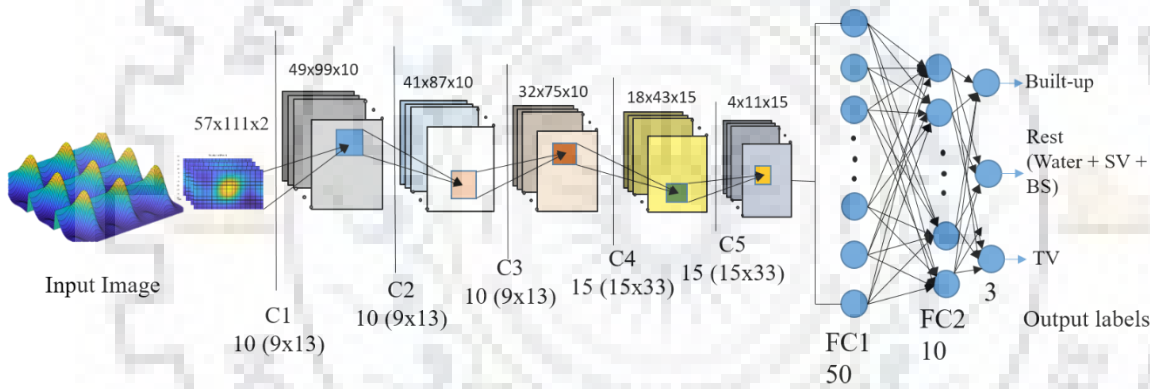
Figure 4.6 Architecture of the proposed PSs-CNN.

Table 4.3 Polarization signature statistics based on window sixes. The values are computed over class-wise ground truth samples.

Class	Window (7X7)		Window (3X3)		Pixel-wise	
	Peak (Mean)	Peak (Range)	Peak (Mean)	Peak (Range)	Peak (Mean)	Peak (Range)
			(in 10^{10})			
Water	3	1-8	3	2-3.5	2.7	0.46 -21
BS	5.5	3-8	4.9	2-10	4.9	1.1-20
SV	6	2-12	6	4.5-7	6	1.8-15
Built-up	17	10-25	19	15-35	19	3.6-37
TV	22	15-30	21	10-35	21	9.6-40

The developed PSs-CNN model is trained on labelled samples collected via ground survey (refer Table 4.1). During the training of PSs-CNN, the bare soil, SV, and water class samples are combined and trained as one output class i.e." rest" in order to artificially create a flood like scenario. The loss function considered during training is the categorical cross entropy function. The learning rate is set to 0.01 and the number of training epochs is set to 200. Adam optimizer is used for network weight optimization. Model architecture and hyperparameters are

88

set similar for Roorkee and Haridwar datasets for consistency purposes and are listed in Table 4.2.

Figure 4.7(a) shows the classified image for Roorkee and Figure 4.7(c) shows the classified image for Haridwar using the PSs-CNN model as classifier. To assess the classification performance of novel PSs-CNN model, the model performance is compared with two classifiers that are mostly used with PSs in land cover classification i.e. the normalized signature correlation mapper (NSCM) based minimum distance classifier and the normalized Euclidean distance (NED) based minimum distance classifier.

Signature correlation mapper or SCM is a cross correlogram constructed by calculating the cross-correlation between two polarimetric signatures. The SCM obtains from the Pearson correlation coefficient and varies from -1 to 1. The SCM between two signatures p_1 and p_2 is given by equation (4.1).

$$SCM(p_1, p_2) = \left(\frac{m\sum_1^m p_1 p_2 - \sum_1^m p_1 \sum_1^m p_2}{\sqrt{\left[m\sum_1^m p_1^2 - \left(\sum_1^m p_1 \right)^2 \right]}\sqrt{\left[m\sum_1^m p_2^2 - \left(\sum_1^m p_2 \right)^2 \right]}} \right) \tag{4.1}$$

where m is the number of polarization states [149], [194]. NSCM is obtained by normalizing the range of SCM between 0 and 1 using equation (4.2).

$$\text{NSCM} = 0.5 \times (\text{SCM} + 1) \tag{4.2}$$

Euclidean distance or ED is the distance metric which provides a quantitative measure of the distance between vectors representing the two polarimetric signatures, and it reflects the dissimilarity between the polarimetric signatures [149], [196]. The ED between two signatures p_1 and p_2 is given by equation (4.3).

$$ED(p_1, p_2) = \sqrt{\sum_1^m (p_1 - p_2)^2} \tag{4.3}$$

where m is the number of polarization states. Further, NED is the normalized version of ED. Both NSCM and NED have been utilized with PSs for land cover characterization in [149] and for land cover classification in [193].

Figure 4.7(e) and Figure 4.7(g) shows classified images using the NSCM classifier (minimum distance is considered as classifier) for Roorkee and Haridwar respectively whereas Figure 4.7(i) and Figure 4.7(k) shows the classified images using the NED classifier (minimum

distance is considered as classifier) for Roorkee and Haridwar respectively. It is evident form Figure 4.1, Figure 4.2, and Figure 4.7(a, c, e, g, i, and k) that the PSs-CNN outperforms the NSCM and the NED on both, the Roorkee, and the Haridwar datasets. Because of the point-wise operability, the NED and NSCM classifiers are not able to capture and exploit the neighbourhood relationship of a PS pixel and thereby are not able to efficiently segregate the built-up areas as discussed in section 4.1. The developed PSs-CNN classifier resolves this issue significantly by extracting and utilizing the subtle differences in the PSs of built-up and TV and successfully separates the two. It is evident from Figure 4.7(a) & Figure 4.7(c) that PSs-CNN classifier is successful in segregating the built-up (represented by red colour) and the TV (represented by green colour) classes. Built-up class is extracted from the PSs-CNN, NSCM, and NED classified images are show in Figure 4.7(b, f, and j) and Figure 4.7(d, h, and l) for Roorkee and Haridwar respectively. It is evident that the PSs-CNN is able to extract the built-up areas (highlighted in white) very efficiently.

Table 4.4 summarizes the quantitative assessment of classification performance of the PSs-CNN, NSCM, and NED classifiers for Roorkee and Haridwar datasets. The PSs-CNN classifier performs land cover classification with good overall accuracy (OA), i.e. 87% for Roorkee and 89% for Haridwar dataset. In contrast, the NED performed the classification with 75 % OA for Roorkee and 73% OA for Haridwar, and the NSCM performed the classification with 74% OA for Roorkee and 76% OA for Haridwar. Both NED and NSCM show low kappa values (refer Table 4.4) whereas the kappa for PSs-CNN is relatively high. Furthermore, the PSs-CNN classifier successfully segregates the built-up class. The built-up class is segregated with 81.8% user and 90.0% producer accuracy for Roorkee and 82.5% user and 92.5% producer accuracy for Haridwar using the PSs-CNN classifier which supports the visual interpretations.

(a)   (b)   (c)   (d)

(e)   (f)   (g)   (h)

(i)   (j)   (k)   (l)
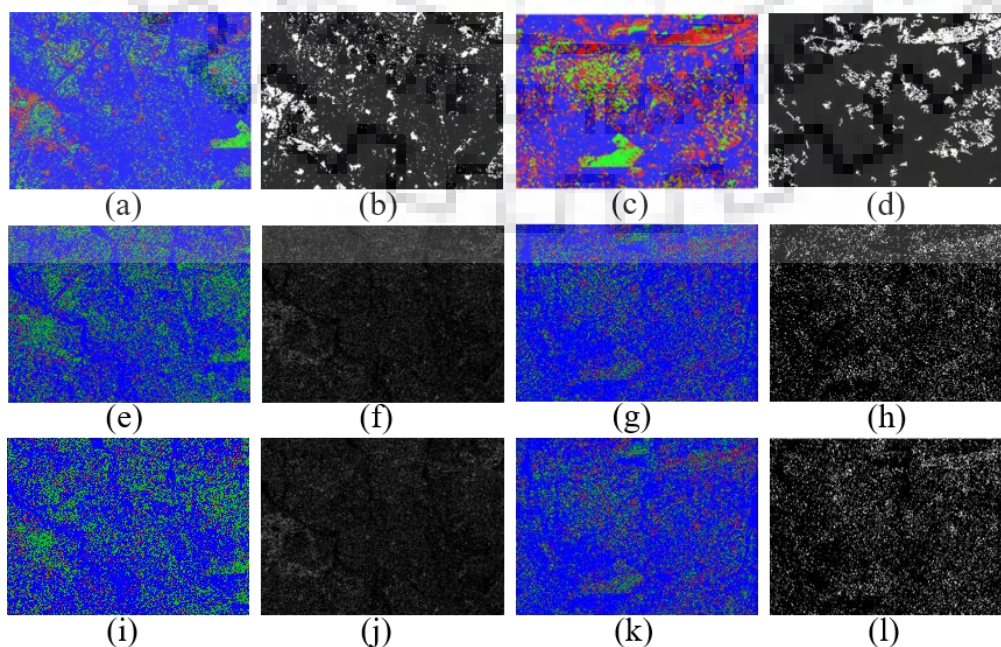
Figure 4.7 Roorkee classified image obtained using: (a) the PSs-CNN classifier, (e) the NED classifier, and (i) the NSCM classifier. Areas classified as built-up in Roorkee study area using: (b) the PSs-CNN classifier, (f) the NED classifier, and (j) the NSCM classifier. Haridwar classified image obtained using: (c) the PSs-CNN classifier, (g) the NED classifier, and (k) the NSCM classifier. Areas classified as built-up in Haridwar study area using: (d) the PSs-CNN classifier, (h) the NED classifier, and (l) the NSCM classifier. Colour scheme: red = built-up, green = TV, and blue = rest (Water + BS + SV).

Table 4.4 Summary of the classification performance of the considered classifiers considered for Roorkee and Haridwar datasets. OA = overall accuracy. Rest = BS + Water + SV.

Dataset	Method	OA (%)	Quadratic Kappa	User accuracy (%)			Producer accuracy (%)		
				Rest	TV	Built-up	Rest	TV	Built-up
Roorkee	PSs-CNN	87.0	0.77	96.2	71.4	81.8	85.8	87.5	90.0
	NSCM	74.7	0.56	86.7	57.6	60.6	83	75	50
	NED	75.0	0.55	86.2	56.8	63.6	83.3	72.5	52.5
Haridwar	PSs-CNN	89.0	0.80	96.5	75.8	82.5	93.3	72.5	92.5
	NSCM	76.0	0.58	91.8	52.7	58.4	85	47.5	77.5
	NED	73.0	0.51	83.7	51.2	62.1	83.3	47.5	67.5

4.3.3. Detection and mapping of flood affected villages using *apriori* *knowledge and image morphology*

This section presents post classification methodologies based on *apriori* knowledge and image morphology to map the villages affected by flood from built-up area image obtained in stage 1. This section covers stage 2 and stage 3 of the proposed approach. The methodology presented in stage 2 detects villages in the built-up area image obtained in stage 1 (refer section 4.3.2). The methodology utilizes *apriori* knowledge and image morphology to build village detection filters. These filters detect villages in the built-up area image and significantly remove false alarms or misclassifications. Development of this methodology is discussed under stage 2 in section 4.3.3.1. Stage 3 presents a novel methodology to identify and map villages that are affected by flood and mask out villages that are not affected by flood. This methodology also utilizes image morphology to map flood affected villages. Development of this methodology is discussed in section 4.3.3.2.

4.3.3.1. Stage 2: Village detection

The focus of this stage is to detect villages in the built-up area image obtained in stage 1 in section 4.3.2. In the methodology, at first, the built-up area image —a binary image— is cleaned before detection because the image might contain objects (built-up areas) that are close to each other and could be part of same village or, there might be objects with holes in them which is undesirable. These artefacts could lead to inaccurate detection and mapping of villages.

Therefore, a series of morphological operations are utilized to clean the binary image. The objects or components which are close to each other, are merged together at first using a morphological image *closing* filter provided in equation (4.4) [197].

$$I_{closed} = (I \oplus S) \Theta S \tag{4.4}$$

Where, I_{closed} is the closed image, I is the image on which the morphological closing operation is to be performed, S is the structuring element with which the closing operation is performed, $\oplus$ and Θ represents the morphological dilation and erosion operations respectively. The vicinity, within which two objects can be merged together is set via the size of the structuring element parameter. The vicinity is set based on *apriori* knowledge of the approximate minimum distance between two villages and image pixel resolution.

The approximate minimum distance between two villages is considered here to 200 meters (based on ground surveys and experiment) and the ALOS PALSAR-2 PolSAR data considered in this study has 6-meter spatial resolution. Therefore, S is set as a disc of 15 pixels' diameter (100 meters) and two objects within this distance are merged as one. The effect of the closing operation is highlighted in Figure 4.8. A segment of built-up area image shown in Figure 4.7 is shown in Figure 4.8(a). Figure 4.8(b) depicts the effects of closing operation on the segment. It can be observed from Figure 4.8(b) that some objects are merged together.

Next, the holes in the objects in the I_{closed} image are to be filled. The holes are filled using equation (4.5) [197].

$$I_{filled} = I_{closed} \bigcup X_k \tag{4.5}$$

Where, I_{filled} is the image with holes filled, I_{closed} is the image with holes obtained from equation (4.4), $\bigcup$ is the *union* operator, and X_k is obtained through iterative process governed by equation (4.6) [197],

$$X_k = (X_{k-1} \oplus S) \bigcap I_{complement} \tag{4.6}$$

Where again, X_0 is the I_{closed} image, S is the connectivity, and $I_{complement}$ is the negative of I_{closed}. Figure 4.8(c) depicts the result of the hole filling operation.

Finally, an area limit is set to filter out objects that are not villages. The idea is that any object considered to be a village is to have at least 5000 square meters of ground area. The area limit is established based on *apriori* knowledge of the study region. An area limit (AL) of 5000

square meters (140 pixels approx. in case of PALSAR-2 image considered here) is set. Equation (4.7) demonstrates the process of filtering objects based on its area.

$$(O_k)_{I_{filtered}} = \begin{cases} 1 & \text{if } (O_K)_{I_{filled}} > \text{AL} \\ 0 & \text{if } (O_K)_{I_{filled}} < \text{AL} \end{cases} \text{ for } k = 1, 2, ..N \tag{4.7}$$

Where, $(O_k)_{I_{filtered}}$ is the kth object in the filtered image and $(O_k)_{I_{filled}}$ is the kth object in the I_{filled} image, AL is the area limit, and N is the total number of objects in I_{filled} image.

Figure 4.8(d) depicts output of the area based filtering. It is observed from Figure 4.8(d) that many unwanted objects have been removed successfully. The objects remained in the filtered image are successfully detected as villages.

The closing, hole filling, and the area filter operations in sequence act as a 'detection filter' which detects villages in a built-up area image and mask out false alarms or misclassifications. This village detection filter is applied to the built-up area images of the PSs-CNN, NSCM, and NED.

Output images after applying the village detection filter are shown in Figure 4.9. These output image(s) which contains only the successfully detected villages is termed as "DV" image(s) hereafter in this chapter.

The DV images corresponding to the PSs-CNN, NSCM, and NED classifiers are assessed using reference data. The reference images for both Roorkee and Haridwar are provided in section 4.2.3 (refer Figure 4.3). The reference images contain all the villages in the corresponding study areas. Figure 4.9 shows false colour composites (FCC) of the DV images of the PSs-CNN, NSCM, and NED, and reference images for Roorkee and Haridwar areas. Figure 4.9a) shows the FCC namely reference-$DV_{PSs-CNN}$-DV_{NED} for Roorkee. In Figure 4.9(a), red colour objects are reference objects (actual villages), green colour objects are villages as per the $DV_{PSs-CNN}$, and blue colour objects are villages as per the DV_{NED}. Further, white colour objects are objects which are present in all three images i.e. the reference image, the $DV_{PSs-CNN}$ image, and the DV_{NED} image.

It means that the object is actually a village and is correctly extracted by both, the PSs-CNN and the NED. Yellow colour objects are villages as per the reference and the PSs-CNN but not as per the NED. Cyan colour objects are villages as per the reference and the NED but not as per the PSs-CNN. The abundance of yellow colour in Figure 4.9(a) indicates to the efficient extraction of built-up areas by the PSs-CNN classifier. Similarly, the

reference-$DV_{PSs-CNN}$-DV_{NSCM} false colour composite for Roorkee shown in Figure 4.9(c) also evidences the better performance of the proposed PSs-CNN classifier.

However, the reference-$DV_{PSs-CNN}$-DV_{NED} and the reference-$DV_{PSs-CNN}$-DV_{NSCM} false colour composites for Haridwar in Figure 4.9(b) and Figure 4.9(d) show numerous green colour objects. These objects are falsely detected as villages as per PSs-CNN and indicate false alarms. A quantitative analysis is therefore required to assess the performance of the classifiers in more detail.
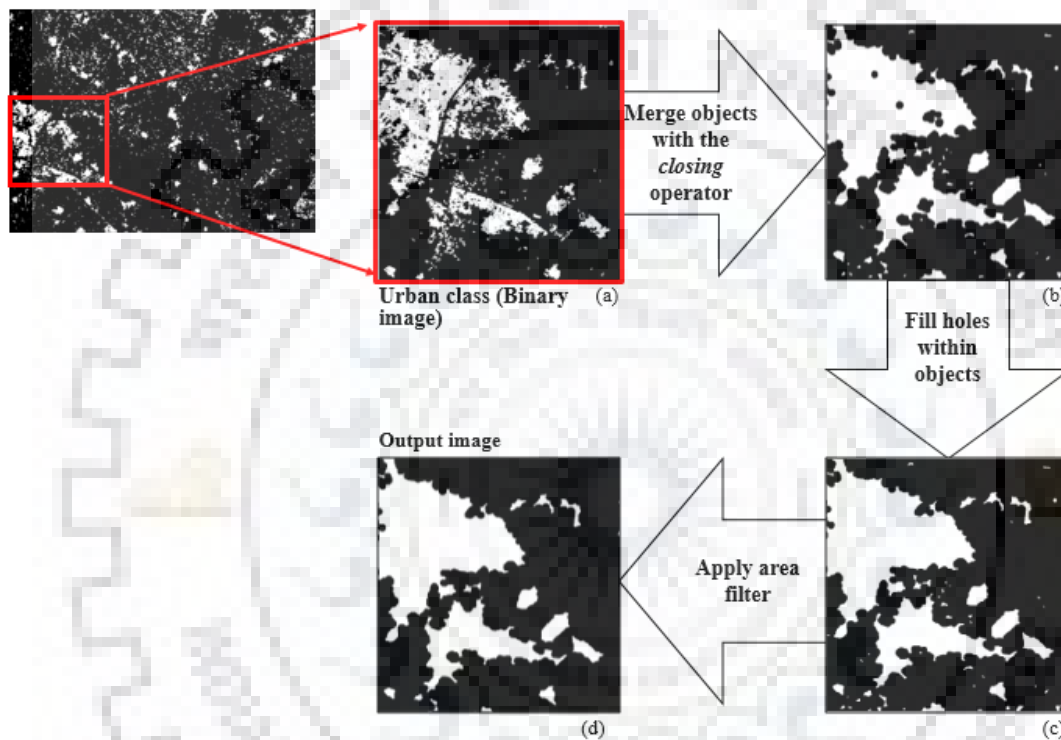
Figure 4.8 Results from different morphological operations considered during village detection: (a) Segment of image shown in Figure 4.7(a). Corresponding segments of: (b) Image after the *closing* operation, (c) Image after the *holes filling* operation, and (d) image after the *filtering* operation.
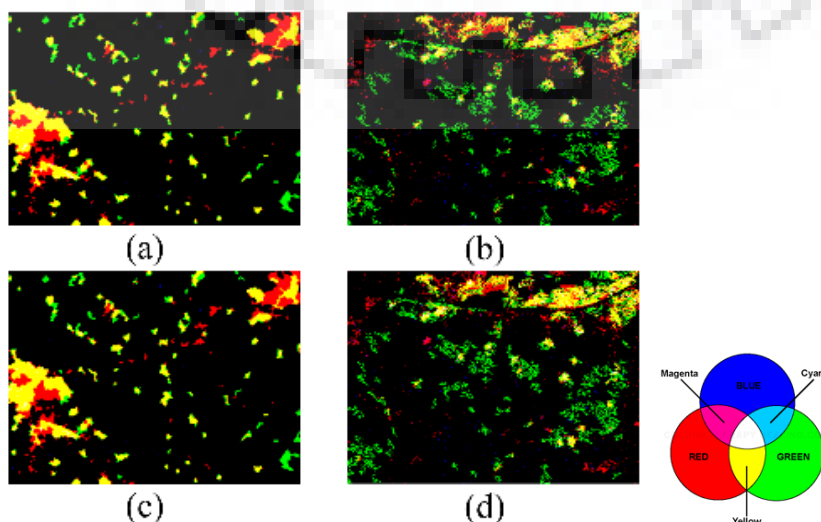
Figure 4.9 (a) False colour composites for Roorkee using reference-$DV_{PSs-CNN}$-DV_{NED} , (b) False colour composites for Haridwar using reference-$DV_{PSs-CNN}$-DV_{NED} , (c) False colour composites for Roorkee using reference-$DV_{PSs-CNN}$-DV_{NSCM} , and (d) False colour composites for Haridwar using reference-$DV_{PSs-CNN}$-DV_{NSCM} . Colour scheme: red = reference, green = $DV_{PSs-CNN}$, blue = DV_{NED} or DV_{NSCM}

For quantitative evaluation of the DV images corresponding to the PSs-CNN, NSCM, and NED classifiers, at first, a bounding box to mark the extent of each object in both the reference and the test image is computed. Then, a bounding box overlap ratio (BBOR) is computed using equation (4.8) for each object pair of the reference and the test image.

$$BBOR = \frac{Overlap\ Area_{(test,reference)}}{min(Bounding\ Box\ Area_{refrence}, Bounding\ Box\ Area_{test})} \tag{4.8}$$

If the BBOR of a particular object pair is above a predefined threshold, then the corresponding object in the test image is counted a successful detection, otherwise not. The threshold is set to 0.1 based on experiment and critical analysis, and is set identical for both Roorkee and Haridwar study areas for consistency purpose.

True positives (villages successfully detected), true negatives (villages missed), and false alarms or false positives (objects detected as villages in the test image but are actually not) are computed using equation (4.9, 4.11, and 4.12).

$$True\ Positive\ (TP) = \sum_{i=1}^{m} logical(BBOR > threshold) \tag{4.9}$$

Where, m is the total number of objects (villages) in the reference image, and

$$logical(x) = \begin{cases} 1 if\ x > threshold \\ 0 if\ x < threshold \end{cases} \tag{4.10}$$

$$True\ Negative\ (TN) = Total\ number\ of\ villages\ in\ reference\ image - TP \tag{4.11}$$

$$False\ Positive\ (FP) = Total\ number\ of\ objects\ in\ test\ image - TP \tag{4.12}$$

True positives (TP), true negatives (TN), and false positives (FP) are computed for Roorkee and Haridwar study areas for PSs-CNN, NED, and NSCM. Table 4.5 summarizes the quantities assessment of the DV images of the PSs-CNN, NED, and NSCM classifiers. The PSs-CNN successfully extracts 75 (out of 93) villages in the Roorkee study area and 47 (out of 60) villages in the Haridwar study area. Whereas, NED successfully extracts 26 villages in Roorkee and 20 villages in Haridwar, and NSCM successfully extracts 22 villages in Roorkee and 27

villages in Haridwar. The low detection rate in case of the NED and the NSCM resonates with their low classification accuracies for built-up class.

The false alarms are also high for the NED and the NSCM. The low values of kappa for NED and NSCM (please refer Table 4.2) indicate to the high false alarms. The PSs-CNN offers higher detection rate and low false alarms for both Roorkee and Haridwar study areas.

Table 4.5 summary of quantitative assessment of performance of PSS-CNN, NED, and NSCM

Dataset	Measures	NED	NSCM	PSs-CNN
Roorkee	Total number of villages in the study area	93	93	93
	TP (Villages successfully mapped)	26	22	75
	TN (Villages missed)	67	71	18
	FP (or False Alarms)	31	41	1
Haridwar	Total number of villages in the study area	60	60	60
	TP (Villages successfully mapped)	20	27	47
	TN (Villages missed)	40	33	13
	FP (or False Alarms)	33	39	5

4.3.3.2. Stage 3: Mapping flood affected villages

Since, a PolSAR image acquired to monitor a flood might include both, flooded and not flooded regions. Therefore, it is necessary for any approach which maps only flooded villages   to mask out the non-flood affected villages present in the PolSAR image. The village detection methodology proposed in section 4.3.3.1 does not separate out flooded and non-flooded villages. Therefore, a new methodology is proposed to achieve the task. This methodology presented in this section encompasses stage 3 of the overall approach.

Intuitively, the villages that are flooded must be completely enclosed by water, otherwise not. This means that any village affected by flood should appear as a 'hole' in the water class (rest class in this case) image and any village that is not affected by flood should not. Based on this observation, a criterion is set i.e. if an object in the DV image obtained in section 4.3.3.1 is a subset of an object in the rest class image, then, that object in the DV image is a flooded village otherwise not. This criterion will separate the flooded village from non-flooded villages.

To realize the criterion, a bounding box overlap ratio (BBOR2) is computed over each pair of DV image object and the "holes filled" version of rest class image object. The holes filled rest class image is obtained using equation similar to (4.8) If the BBOR2 is 1, then the corresponding object in the DV image is a flooded village and if the BBOR2 is less than 1, the corresponding object in the DV image is not a flooded village. This methodology is depicted in Figure 4.10.

The methodology is evaluated on three different scenarios on the DV image of Roorkee corresponding to the PSs-CNN. The three different scenarios are shown in Figure 4.11. In the first scenario i.e. scenario 1, flood has affected the whole study region (already discussed in

section 4.3.2) whereas, in the second and third scenarios i.e. scenario 2 and 3, flood has affected parts of the study region (partial flooding). When the DV image of Roorkee corresponding to the PSs-CNN is subjected to these three scenarios, the resulting flood situations are shown in Figure 4.12(a, b, and c). It is clear from Figure 4.12(a) that all the villages are affected by flood whereas in Figure 4.12(b) and Figure 4.12(c), only few villages are affected. Employing the proposed methodology to the three flood situations results in identification and mapping of flood affected villages. The flood affected villages identified in the three flood situations are highlighted in yellow in Figure 4.12(a, b, and c). It is evident from Figure 4.12(a, b, and c) that the methodology proposed in this section successfully identifies and maps villages that are affected by flood and masks out villages that are not affected by flood.

In summary, two methodologies are proposed in section 4.3.3. The first methodology developed in section 4.3.3.1 detects villages and removes false alarms and misclassifications in a built-up area image obtained from the PSs-CNN classifier. The second methodology developed in section 4.3.3.2 detects maps flooded villages and masks out non-flooded villages. These two methodologies in combination with the PSs-CNN based built-up extraction model completes the flood affected village mapping approach.
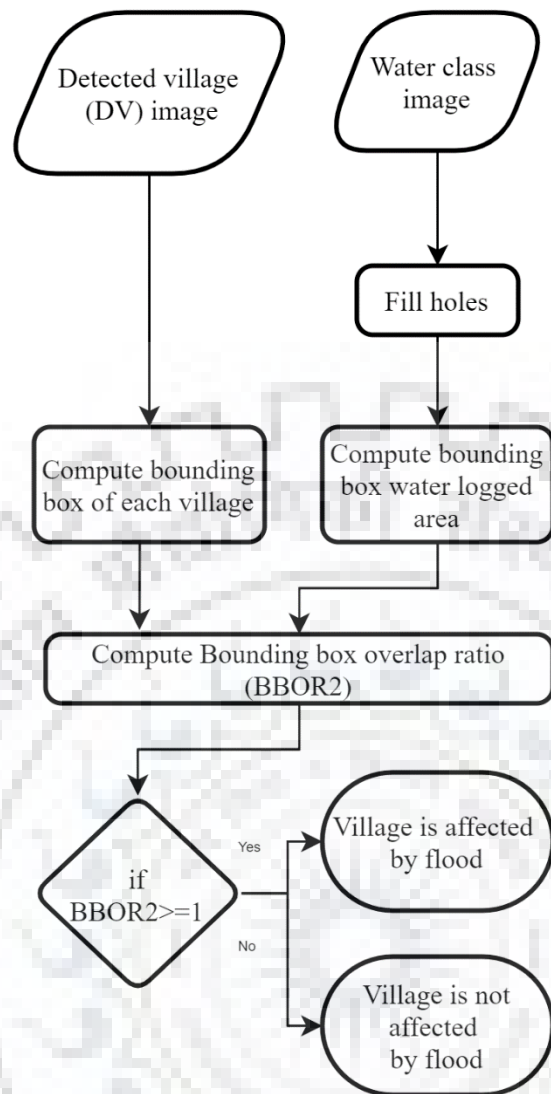
Figure 4.10 Methodology flowchart for flood affected village mapping.

(a)

(b)

(c)
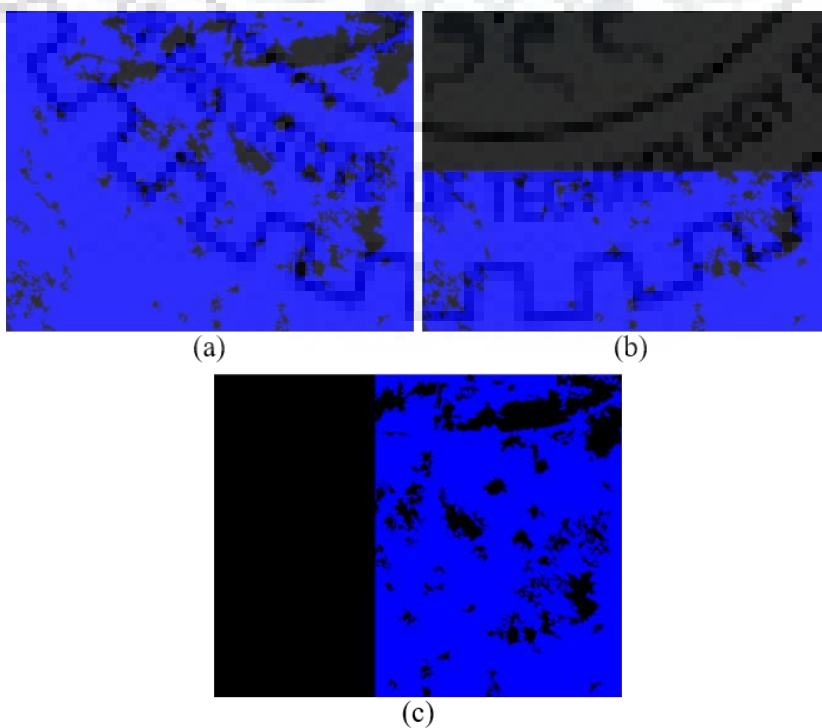
Figure 4.11 Different scenarios: (a) Scenario 1, (b) Scenario 2, and (c) Scenario 3. Colour scheme: blue = flooded region.
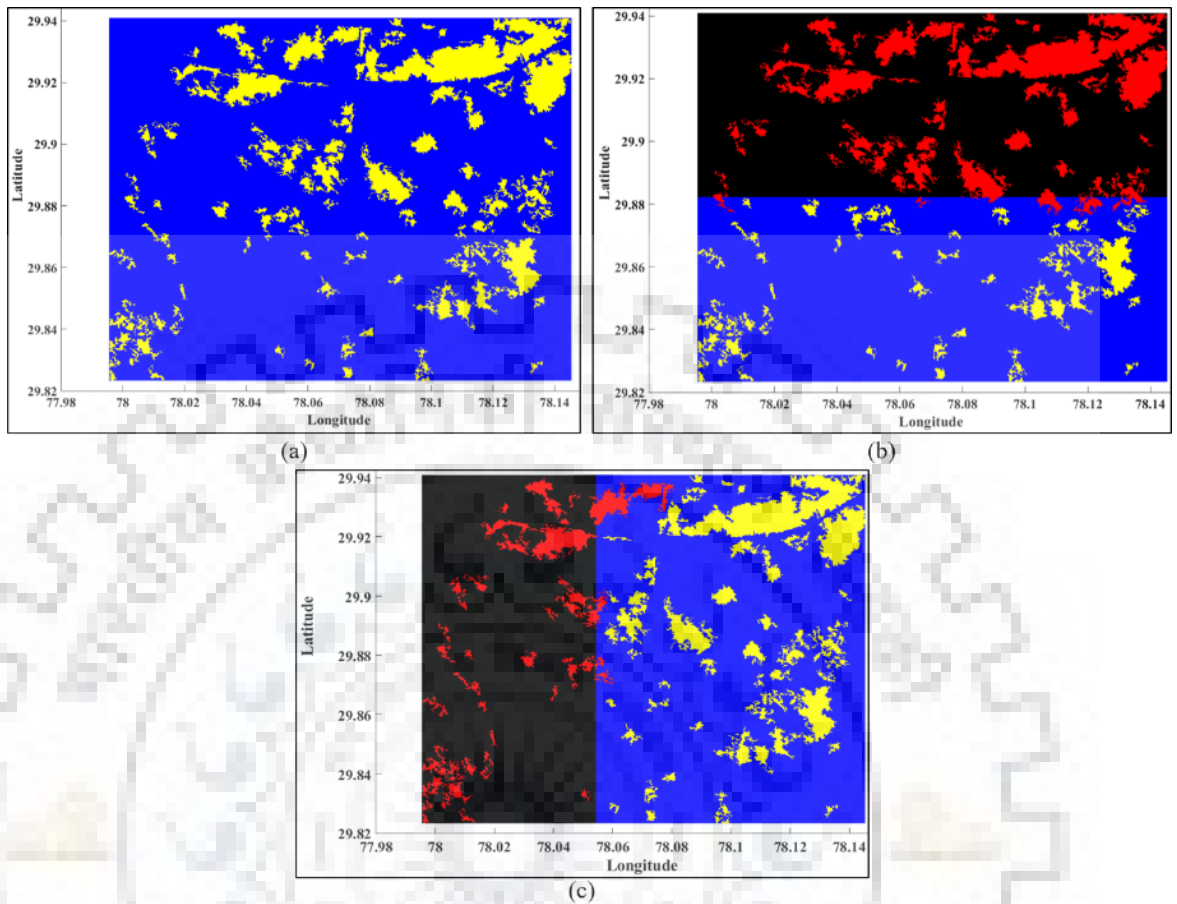
Figure 4.12 Flood maps: (a) Affected villages under flood condition 1. (b) Affected villages under flood condition 2. (c) Affected villages under flood condition 3. Colour scheme: red = villages not affected by flood, yellow = villages affected by flood, and blue = flooded region.

4.4. Conclusion

This chapter presented an efficient approach to map flood affected villages using PolSAR image. The approach included three stages i.e. –*first*, built-up area extraction, –*second*, village detection, and –*third*, flood affected village mapping.

In the first stage, the approach utilized PolSAR polarization signatures (PSs) and CNNs to extract built-up areas from a PolSAR image. A novel PSs-CNN classifier is developed. The PSs-CNN model provided improved built-up area extraction performance in contrast to the NED and the NSCM based classifiers. In the second stage, the approach introduced and implemented a novel village detection methodology. The methodology uses a detection filter to detect whether a particular object in the built-up area image (obtained in stage 1) is a village or not. The detection filter uses *apriori* knowledge and image morphology to detect villages in the built-up area image. The detection methodology also significantly removes false alarms and misclassifications. In the

third and final stage, flood affected villages are identified and non-flooded villages are masked out using another image morphology based methodology. This methodology computed overlap ratios between objects (successfully detected as villages) and flooded area and utilized it for identification of flood affected villages.

Quantitative and qualitative assessment of the performance of the novel approach over two different study areas is carried out. Different artificial flood scenarios are created to evaluate the performance of developed approach. The developed approach is evaluated under these flood scenarios and flood affected villages are identified and mapped. Number of villages successfully detected and false alarms are computed with the help of ground truth information (actual number of villages or towns in the study regions are known). Results indicate to high detection accuracy (above 90%). The approach may be scalable and potentially satisfactory for real time usage.

PERMUTED SPECTRAL AND PERMUTED SPECTRAL-SPATIAL CNN MODELS FOR POLSAR-MULTISPECTRAL DATA BASED LAND COVER CLASSIFICATION

In contrast to chapter 3 and chapter 4 where novel DNN models are proposed and developed for unimodal satellite image, this chapter explores the performance of conventional DNN models with multisensor/multimodal satellite images. Also, based on the performance of these conventional models, novel DNN models are proposed here.

In context to satellite based multisensor earth observation data, today, both synthetic aperture radar (SAR) and multispectral (MS) image data are easily available with good spatial and temporal resolutions. The two data modalities with different spectral characteristics complement each other in many earth observation data applications such as land cover classification, crop monitoring, and change detection. However, it is still a challenge to develop techniques, methods, or approaches which can process the two data modalities together efficiently and without losing information from either. This chapter presents a study which attempts to introduce deep learning and DNN based novel models for multisensor (, PolSAR-MS) satellite image data processing that utilizes localized convolutions in both spatial and spectral domains to perform land cover (LC) classification. The study also attempts to establish the significance of performing ensembled localized spectral convolutions in comparison to non-localized or localized spectral convolutions. Two novel models namely; permuted local spectral convolutional neural network (Perm-LS-CNN), and permuted local spectral-spatial convolutional neural network (Perm-LSS-CNN) are proposed in order to achieve better classification results in terms of both accuracy and generalization.

"Generalization is the measure of the ability of machine learning techniques on how well these techniques perform on 'out of sample' data."

5.1. Introduction

The synthetic aperture radar (SAR) and multispectral (MS) image data modalities are crucial for earth studies and complement each other in many earth observation applications such as land cover classification, crop monitoring, or change detection. In general, multispectral image provides spectral reflectance information and PolSAR image provides scattering and polarizing behaviour information about the earth surface objects. These earth surface objects are more commonly termed as land covers. The basic idea behind utilizing the two image data modalities together is to acquire better spatial, spectral, and temporal information on the land covers. Although, higher spectral information can be achieved through hyperspectral images but at a cost of spatial and temporal information. Also, since it is an earth observation strategy based on optical imaging, therefore it suffers from the classical challenges (clouds, night-time). Similarly, use of polarimetric SAR (PolSAR) in standalone fashion for land cover classification resulted in a maximum of 75% classification accuracy in mixed land cover class scenarios (refer section 3.7). For applications such as land use land cover classification or land cover monitoring, all three aspects are crucial and a multisensor data based approach is more beneficial. Using PolSAR and MS image data modalities together is shown to be a convenient and successful strategy. Methodologies listed in many articles such as in [52], or more recently in [44], [55], [198] have been used by the earth observation data processing community in processing multispectral and PolSAR data together. Authors in [44] provide a meta-analysis on the methodologies in successful applications of PolSAR and MS image data fusion. The methodologies include traditional, machine learning, and knowledge based decision trees. In [44], authors also remarked that the studies in multisensor data fusion based land cover classification mostly use MS and PolSAR data with machine learning methods. Most popular machine learning based multisensor data fusion techniques include artificial neural networks (ANNs or NNs), and support vector machines (SVMs). In fact, studies in [44]–[46] and [55] have also recognized the significance of machine learning in various multisensor image data based applications. From the aforementioned review, two points are evident; -first, utilization of PolSAR and MS data together is a successful strategy, and -second, machine learning techniques are successful in utilizing the two i.e. PolSAR and MS, image data modalities together.

However, it is challenging for (NNs) with shallow depth (number of layers ≤ 2) to extract high level abstract/complex features from high dimensional earth observation image data without overfitting or losing generalization [199]. Generalization refers to the ability of how well an algorithm performs on 'out of the seen samples' data points. If an algorithm is performing well

outside training data, then the algorithm is showing good generalization performance and vice versa. Researchers in many studies[5], [58], [100] are exploring new advanced machine learning approaches that are able to extract complex features from high dimensional earth observation image data. In the direction, deep neural networks with deep learning are the potential and powerful tools for extracting high level information from earth observation image data[126], [200] Convolutional neural networks (CNNs) are recently the most widely used deep network models in earth observation image data based land cover classification [58]. CNNs have the ability to extract complex features directly from data [35], [201], [202]. Another advantage of developing CNNs for PolSAR-MS multisensor image data processing is that CNNs can directly utilize multisensor high dimensional earth observation image data with efficient performance. Unlike the conventional data fusion methods, this reduces the need of fusing the PolSAR-MS images together before using it further.

In [5], authors present a comparison of CNN based models with few popular earth observation image data classifiers in processing hyperspectral images. The CNN model in [5], is designed to effectively capture global spectral features along with local spatial features. Similar studies are reported in [96],[101] and [203]. The CNN models in these studies employed one, two, and three dimensional convolutions to extract spectral and spatial information from unimodal earth observation image data. Moreover, CNNs have also been used with multisensor image data. For example, CNNs have been utilized with PolSAR-MS data in earth observation image data applications. In [72], authors used one and two dimensional CNNs on Landsat8-Sentinel1 multisensor image data for land cover and crop type classification and further compared the results with random forest (RF) classifier. Their model reports better classification performance by capturing local contextual (spectral and/or spatial) features whereas the RF classifier captured only the global dominant features. The study in [72], indicates potential of CNNs with PolSAR-MS multisensor image data for earth observation applications and motivates for similar studies.

However, one important aspect which comes to consideration while processing this PolSAR-MS data using CNNs is the efficient exploitation of the spectral information present in the high dimensional multisensor (PolSAR-MS) data. The exploitation of spectral information is crucial for CNN based multisensor data applications. Exploitation of spectral information present in hyperspectral image has been done in past, either exclusively with one dimensional CNNs or, along with the spatial information using two, or three dimensional CNNs. Many studies [100], [101], [204]–[206] and more have used one dimensional CNNs as spectral feature extractors for hyperspectral image based classification. In contrast, extracting spectral information along with spatial information from hyperspectral image using CNNs is a recently

popular approach. The approach has however improved the results in land cover monitoring and classification applications. Because hyperspectral images have large spectral dimensions, extracting spectral information is computationally expensive and compounding with extraction of spatial information makes it worse. Therefore, some studies [5], [102], [105], [207] used principal component analysis (PCA) to reduce the high dimensionality of hyperspectral image and then used two dimensional CNNS for classification applications. While, alternate to this, some studies [108], [207], [208] used one and two dimensional CNNs separately to extract the spectral and spatial information and combined the separately extracted features later. Further, some studies used three dimensional CNNs to extract local spectral and spatial information simultaneously. In [96], [98], authors used three dimensional CNNs as spectral-spatial feature extractors and classifiers for hyperspectral satellite image data.

In summary, enough evidence suggests that; one, two, and three dimensional CNNs have been successfully used to extract spectral and/or spatial information from hyperspectral image data. However, the significance of spectral and/or spatial CNNs as land cover classifiers with high dimensional multisensor image data is still interesting to investigate. Therefore, the study presented in this chapter focusses on two aspects.

- To critically analyse the one dimensional (1D), two dimensional (2D), and three dimensional (3D) CNN capabilities to perform land cover classification using multi-sensor (PolSAR-MS) data.
- Development of novel classification frameworks based on 1D (spectral) or 3D (spectral-spatial) CNNs; in order to improve the generalization ability of the classifier during land cover classification with PolSAR-MS satellite data.

1D, 2D, and 3D CNN frameworks for PolSAR-MS data based land cover classification are discussed in section 5.4. The idea of extracting the spectral information efficiently from PolSAR-MS data using CNNs is achieved by performing a permuted localized spectral convolution alone, and along with localized spatial convolution on PolSAR-MS data. Spectral permutation before localized spectral convolution is proposed to increase the generalization ability of the CNN classifier. Based on this idea, two models are proposed, namely permuted local spectral CNN (Perm-LS-CNN) and permuted local spectral-spatial CNN (Perm-LSS-CNN). Both models are discussed in section 5.5.

The remaining of the chapter is organized as follows; Section 5.2 provides a brief description on popular CNN models for earth observation image data processing. Section 5.3 provides information on the experimental setup i.e. study areas, datasets, ground truth, and other resources considered and used during the study. Section 5.4 discusses the development of conventional 1D, 2D, and the 3D CNN models for processing PolSAR-MS image data and presents internal configurations of these models. Section 5.5 presents two novel CNN based PolSAR-MS data processing frameworks for land cover classification. Results from the

developed methods and compared classifiers are discussed in section 5.6. Finally, section 5.7 concludes the chapter.

5.2. Theoretical Background

5.2.1. Convolutional neural networks

The fundamental concepts on CNNs are discussed in section 3.2.2 with an example of 2D CNNs. Similar networks can be realized for 1D or 3D CNNs accordingly. However, the architectures of 1D, 2D, and 3D CNNs for remote satellite image data processing are as follows.

- For 1D CNNs, the satellite image data input is a vector (usually a vector of spectral features). $\mathbf{W_m}$ is a matrix and b_m is vector. In general, 1D CNNs are used as feature extractors in satellite image data, especially hyperspectral data processing.
- For 2D CNNs, the convolution is performed on the two spatial dimensions (e.g. images). The input may contain extra dimensions (such as spectral information), but these are not used in the convolution. $\mathbf{W_m}$ is a 3D matrix and b_m is vector. It is important to note that in 2D CNNs, local convolutions are only performed in the spatial dimensions and not in the spectral dimension. 2D CNNs are used as spatial feature extractors and classifiers.
- In 3D CNNs, convolutions are performed in the spatial dimensions and the temporal dimension simultaneously. However, in the absence of temporal information and presence of high spectral information, spectral dimension replaces the temporal dimension and local spectral convolutions are possible. $\mathbf{W_m}$ is a 4D matrix and b_m is vector. Therefore, 3D CNNs are used as spatial temporal or spatial spectral feature extractors and classifiers in satellite image data processing.

The present study utilizes and investigates the ability of 1D, 2D, and 3D CNNs as land cover classifiers on PolSAR-MS data. Also, the present study proposes novel LC classification models based on these CNNs.

5.3. Experiment Setup

5.3.1. Study area

Two areas are considered for the analysis, development and testing of the proposed methodologies. The first study area is common from chapter 3 and chapter 4 (Roorkee study area). Details of this study area are presented section 3.3.1. This study area is termed as 'Roorkee study area' in this chapter. The second study area is also common from chapter 4 (Haridwar study area). This second study area is termed as 'Haridwar study area' in this chapter. Details of this study area are presented in section 4.2.1. Both study areas have mixed class type landscape,

106

which is beneficial for the evaluation of classifiers performance in mixed class scenarios. Figure 4.1(a) and Figure 4.2(a) shows a Google Earth images of the Roorkee and Haridwar study areas respectively.

5.3.2. Dataset

Datasets used in the study are multisensor i.e. multispectral (MS) and polarimetric synthetic aperture radar (PolSAR) data acquired over the study areas i.e. Roorkee and Haridwar. Multispectral data used is the Landsat-8 multispectral data acquired on March 17, 2015 (band specifications are provided in Table 1.2. The PolSAR data used is the ALOS PALSAR-2 L band fully polarimetric single look complex PolSAR data acquired on March 13, 2015 (common from chapter 3 and chapter 4).

Preprocessing is required with both MS and PolSAR raw data. Multispectral data mostly suffers from atmospheric perturbations whereas PolSAR data suffers from speckle noise. Preprocessing is done separately on MS and PolSAR data. Preprocessing of PolSAR PALSAR-2 data is similar to as explained in chapter 3 (refer section 3.3.4). Derived PolSAR features are also used in this study. Polarization signatures (PSs) and further, polarization signatures correlation features (PSCFs) are computed from the single look complex PolSAR data with procedure similar to as in chapter 3 (refer section 3.4). Based on the analysis of chapter 3 (refer section 3.4), four PSCFs namely "corr_co_Di", "corr_co_FP", "corr_co_HD", and "corr_co_VD" are used as derived PolSAR features in the current study. Overall, three basic (HH, HV, and VV) and four derived (corr_co_Di, corr_co_FP, corr_co_HD, and corr_co_VD) PolSAR feature are used in the current study. Figure 5.1(a) and Figure 5.2(a) depicts a false colour composite from the pre-processed PolSAR image data for Roorkee and Haridwar study areas respectively.

Landsat-8 operational land imager (OLI) multispectral data is atmospherically corrected first. Out of the 11 bands shown in Table 1.2, band 2 (blue) through 7 (short wave infrared) are selected for the study. These selected bands are pan sharpened as explained in [209] and then resampled to the PALSAR-2 data spatial resolution of 6 meters using the "bicubic spline" interpolation method [210]. Figure 5.1b) and Figure 5.2(b) depicts a false colour composite from the pre-processed MS image data for Roorkee and Haridwar study areas respectively. The re-sampled, pre-processed MS bands are co-registered with the pre-processed PolSAR channels and the stacked 13 bands (six from multispectral and seven from PolSAR) are used as input data for further analysis and study. It is clear from Figure 5.1 or Figure 5.2 that PolSAR and MS satellite image modalities complement each other. For example, built-up areas and forests in the Roorkee study area exhibit similar backscattering response (light green colour in PolSAR image, Figure

5.1(a)) whereas both the land covers are clearly separable in the MS image (refer Figure 5.1(b)). Similarly, built-up areas are not easily separable from bare soil in MS image whereas the two land covers are easily separable in a PolSAR image.

5.3.3. Ground truth

Five land cover (LC) classes are considered namely bare soil (BS), tall vegetation (TV), short vegetation (SV), built-up, and water. Ground truth class data points are measured directly on the terrain based on visual inspection. Ground truth class data points collected for study in chapter 4 (refer section 4.2.3) are utilized here as well since the class labels are same and the study areas are also same. Details of the ground truth collected for both study areas are shown in Table 5.1. Also, samples images of land covers collected as ground truth are shown in Figure 3.3.
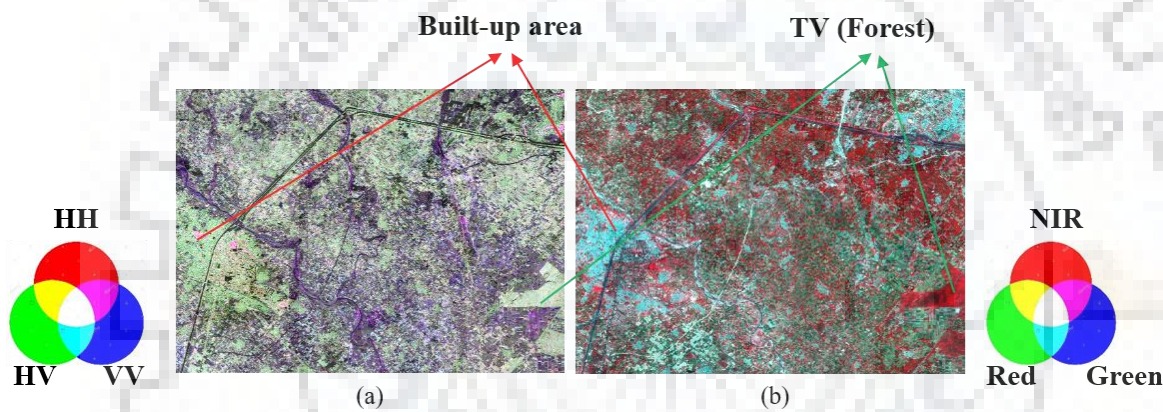
Figure 5.1 Datasets for Roorkee study area: (a) False colour composite (red = HH, green = HV, and blue = VV) of ALOS PALSAR-2 PolSAR data, and (b) False colour composite (red = NIR, green = Red, and blue = Green) of Lansat-8 multispectral data.
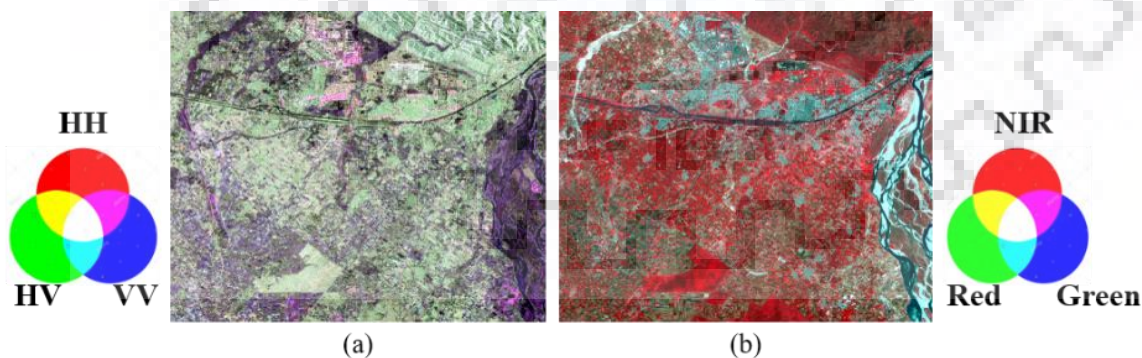
Figure 5.2 Datasets for Haridwar study area: (a) False colour composite (red = HH, green = HV, and blue = VV) of ALOS PALSAR-2 PolSAR data, and (b) False colour composite (red = NIR, green = Red, and blue = Green) of Lansat-8 multispectral data.
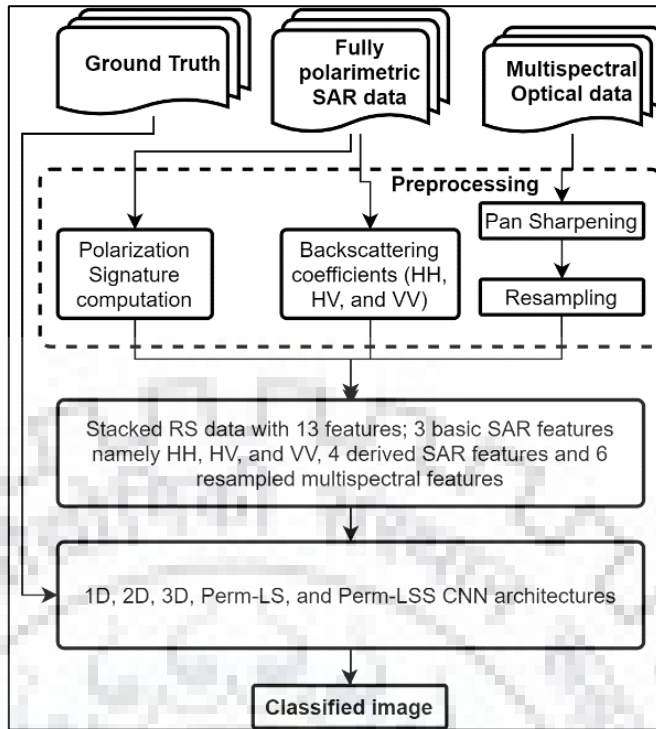
Figure 5.3 Flowchart of the experiment setup and CNN based land cover classification methodologies proposed with PolSAR-MS data

Table 5.1 Summary of the ground truth collected for study. *BS = Bare soil, TV = Tall vegetation, SV = Short vegetation. All points are acquired directly on the field by visual inspection at selected GPS coordinates.

Land Cover	Roorkee		Haridwar	
	Training and validation	Testing	Training and validation	Testing
BS	200	40	200	40
TV	200	40	200	40
SV	200	40	200	40
Built-up	200	40	200	40
Water	200	40	200	40
Total	1000	200	1000	200

The experiment setup i.e. the study areas, the datasets, or the ground truth data points, is used for, first, the critical analysis of the performance of conventional CNN frameworks, and second, the development of novel CNN classification frameworks for land cover classification. The flowchart shown in Figure 5.3 depicts the experiment setup.

5.4. 1D, 2D, and 3D CNNs for Land Cover Classification with PolSAR-MS Data

5.4.1. 1D-CNNs for PolSAR-MS data.

One dimensional CNNs (1D CNNs) have been successfully used with unimodal satellite image data in the recent years. In [204], authors used stacked auto encoders for classification with hyperspectral satellite image data. They also used PCA for hyperspectral feature reduction before

classification as the number of input features is very high (say 176 for KSC dataset). A range i.e. from 4 to 8 principle components are considered as features during classification. PolSAR-MS data used in the present study only have 13 spectral bands and do not require feature reduction. In [100], authors used different 1D-CNNs for two popular hyperspectral satellite image data (the Salimas Valley and the Indian Pines dataset) classification. In [101], authors used 1D-CNN for hyperspectral satellite image data classification. All these references report that 1D-CNNs are good feature extractors for high-dimensional hyperspectral satellite image data. Most of these studies used some form of feature reduction technique before feeding data to the 1D-CNN in order to reduce the computational complexity of the model. The reduced number of bands are on the same order as the number of bands present in the PolSAR-MS data considered here in this study. This argument justifies the direct use of 1D-CNNs for PolSAR-MS satellite image data.

Two variants of 1D CNNs are considered here. In the first 1D CNN model i.e. 1D-CNN-v1, the kernel size of the filter in the first convolutional layer i.e. C1 layer is equal to the spectral depth of PolSAR-MS data i.e. 13. One dimensional spectral convolution is shown in Figure 5.4(b). In the second 1D CNN model, i.e. 1D-CNN-v2, the kernel size is less than the spectral depth of PolSAR-MS data. One dimensional local spectral convolution is visualized in Figure 5.4(c). Models similar to 1D-CNN-v2 has been successfully tested with hyperspectral data [96], [100] The present study focuses on the analysis of the performance of 1D CNNs; both 1D-CNN-v1 and 1D-CNN-v2 with PolSAR-MS data.

The selected hyper-parameter settings are listed in Table 5.2. Apart from kernel size, the rest of the parameters are similar for both models. Rectified linear unit (ReLU) and Softmax activation functions are used in the intermediate and output layers respectively. A dropout strategy is considered for regularization and dropout fraction values of 0.2 for C1 and 0.1 for C2 and FC layer are set. Pooling is not considered because it hinders the full usability of the spectral variation. The loss function used in both models is the categorical cross entropy function. It optimizes the classification accuracy based on probabilities of each class during classification. The learning rate and the number of epochs are set to 0.01 and 200 respectively.
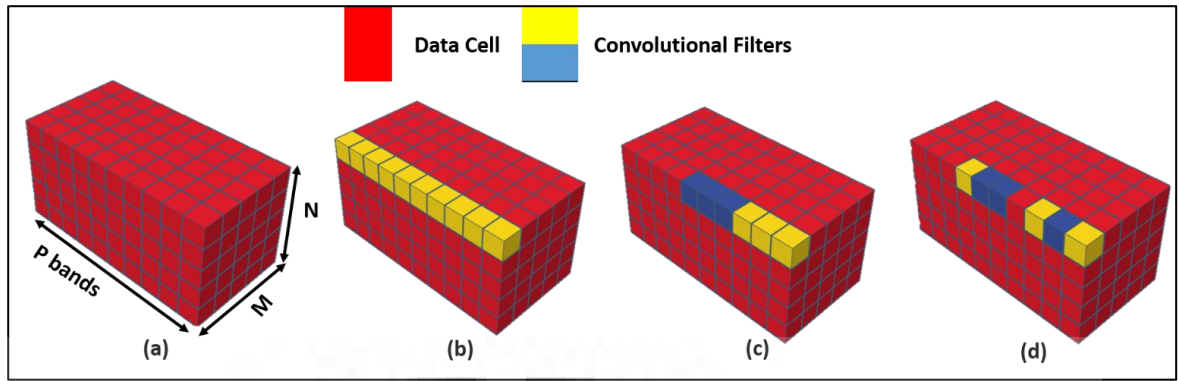
Figure 5.4 Different convolution strategies: (a) Data cuboid with M by N pixels and P bands, (b) Convolution with $1 \times 1 \times P$ convolutional filter where 1×1 represents to the spatial extent and P represents total number of bands in the data. Each filter convolves with all the bands as all times. This strategy is employed in 1D-CNNs during per pixel classification of satellite images. (c) Convolution with $1 \times 1 \times 3$ convolutional filters where 1×1 represents the spatial extent and 3 represents the number of adjacent bands to be convolved simultaneously. One filter can convolve with three adjacent bands only. This strategy is also employed in 1D-CNNs. (d) Convolution with $1 \times 1 \times 3$ convolutional filters where 1×1 represents to the spatial extent and 3 represents number of arbitrarily selected bands to be involved. This strategy is employed in the proposed Perm-LS-CNN framework.

5.4.2. 2D-CNNs for PolSAR-MS data

The use of spectral information without spatial information by 1D-CNNs undermines the full potential of satellite image data. Adding spatial information provides contextual aid to the classifiers, especially CNNs in the classification process. Two dimensional or 2D-CNNs have also been successfully used with unimodal satellite image data in the recent years. In [112], authors used patch based 2D-CNN model for land cover classification with multispectral data. They opted a $5 \times 5 \times 8$ Landsat-8 image patch and a 3×3 2D convolutional kernel. In [82], authors used 2D-CNNs for land cover classification with PolSAR data. Many studies used 2D-CNNs for classification with hyperspectral satellite image data (refer section 2.1). The need for dimension reduction in these studies is already discussed in section 5.4.1. These studies support the utilization of 2D-CNNs with PolSAR-MS data for land cover classification. In [72], authors used 2D-CNNs for crop type classification with PolSAR and multispectral data. They used Landsat-8 multispectral and Sentinel-1 C-Band PolSAR data. Although in [72] authors have studied the PolSAR-MS data potential for crop type classification with 2D-CNNs, the current study differs on: – first, using PolSAR-MS on a single time stamp with a focus on spectral and spatial features. In [72], authors used time series PolSAR-MS data with a focus on temporal and spectral features; and – second, Quad PolSAR data is considered in the present study in contrast to dual PolSAR used in [72]. The increased channels add complementary information which in turn add an indirect weighting of the spatial features. Analysis of fully polarimetric PolSAR with multispectral data for classification is therefore beneficial.

Table 5.2 *1D-CNN-v1*, *1D-CNN-v2, 2D-CNN-v1*, and *3D-CNN-* v1 CNN model configurations, hyperparameter and training settings.

Model	Layer				
	Input	Convolutional layer 1 (C1)	Convolutional layer 2 (C2)	Fully connected layer (FC)	Output layer
1D-CNN-v1	Input data size $= 1 \times 13$ Normalization = Batch Normalization	Filters = 20 Kernel size = 1×13 Activation = ReLU Dropout fraction = 0.2 Pooling size = 0	Filters = 20 Kernel size = 1×1 Activation = ReLU Dropout fraction = 0.1 Pooling size = 0		
1D-CNN-v2	Input data size $= 1 \times 13$ Normalization = Batch Normalization	Filters = 20 Kernel size = 1×4 Activation = ReLU Dropout fraction = 0.2 Pooling size = 0	Filters = 20 Kernel size = 1×4 Activation = ReLU Dropout fraction = 0.1 Pooling size = 0	Nodes =16 Activation = ReLU Dropout fraction = 0.1	Labels = 5 Activation = Softmax
2D-CNN-v1	Input data size $= 3 \times 3 \times 13$ Normalization = Batch Normalization	Filters = 20 Kernel size = 3×3 Activation = ReLU Dropout fraction = 0.2 Pooling size = 0	Filters = 20 Kernel size = 1×1 Activation = ReLU Dropout fraction = 0.1 Pooling size = 0		
3D-CNN-v1	Input data size $= 3 \times 3 \times 13$ Normalization = Batch Normalization	Filters = 20 Kernel size = $3 \times 3 \times 4$ Activation = ReLU Dropout fraction = 0.2 Pooling size = 0	Filters = 20 Kernel size = $1 \times 1 \times 4$ Activation = ReLU Dropout fraction = 0.1 Pooling size = 0		
Training parameters		Loss function = Categorical cross entropy Learning rate = 0.01 Number of epochs = 200 Optimizer = Adam [211] 10 fold cross validation strategy is employed during training			

Figure 5.5: Different convolution strategies: (a) Data cuboid with M by N pixels and P bands, (b) Convolution with $3 \times 3 \times P$ convolutional filter where 3×3 represents the spatial extent and P represents the total number of bands in the data. Each filter convolves with all the bands. This strategy is employed in 2D CNNs, (c) Convolution with $3 \times 3 \times 3$ convolutional filters where 3×3 represents to the spatial extent and 3 represents number of adjacent bands to be convolved simultaneously. One filter can convolve with three adjacent bands only. This strategy is employed in 3D-CNNs [96], and (d) Convolution with $3 \times 3 \times 3$ convolutional filters where 3×3 represents the spatial extent and 3 represents the number of arbitrarily selected bands to be involved. This strategy is employed in the proposed Perm-LSS-CNN framework.

A 2D-CNN model namely 2D-CNN-v1 is proposed here. 2D-CNN-v1 configuration and hyperparameter settings considered are listed in Table 5.2. The idea of a local spatial convolution is shown in Figure 5.5 (b). The configuration for the intermediate and output layers and learning parameters are the same as in section 5.4.1 and are summarized in Table 5.2.

5.4.3. 3D-CNNs for PolSAR-MS data

The 2D-CNN-v1 model discussed in section 5.4.2 is successful because of its spatial contextual awareness. The localized spatial convolutions provide local connectivity, which controls the impact of neighbouring pixels on the pixel to classify. Parameter sharing between local filters also makes the learning faster and less cumbersome, in contrast to fully connected neural networks. In satellite image data applications, spectral information is as important, if not more, as spatial information. Local connectivity in spectral dimension also provides the same benefits. Advantages of local spectral convolutions and local spatial convolutions together can be exploited using 3D-CNNs. The 3D-CNNs have been successfully used with hyperspectral (unimodal) satellite image data in the recent years. Authors in [96], [108], and [203] used 3D-CNNs for spectral spatial information extraction for fusion and classification applications.

The proposed 3D CNN model 3D-CNN-v1 architecture and hyperparameters considered are listed in Table 5.2.. The C1 layer of 2D-CNN-v1 in section 5.4.2 uses the complete spectral depth, i.e. 13 bands, during convolutions whereas 3D-CNN-v1 uses only a subset during convolutions. The idea of local spectral and local spatial convolution is shown in Figure 5.5 (c).

113

For example, a first filter (in yellow) convolves with bands 1, 2, and 3 whereas a second filter (in blue) convolves with bands 5, 6, and 7. This provides an improvement in classification accuracy and robustness compared to other 3D-CNN models [96], [203] Configuration and learning parameters are the same as in section 5.4.1 and are summarized in Table 5.2.

5.5. Proposed Permuted Local Spectral CNN (Perm-LS-CNN) and Permuted Local Spectral-Spatial CNN (Perm-LSS-CNN)

In *1D-CNN-v2*, a local convolution is performed over the spectral dimension. The convolution filter size in the spectral dimension i.e. 4 is less than the total number of input bands i.e. 13. This local spectral convolution improves the classification performance as discussed in [96]. A pictorial representation of local spectral convolution is depicted in Figure 5.4(c). For example, a first filter (in yellow) convolves with bands 1, 2, and 3 whereas a second filter (in blue) convolves with bands 4, 5, and 6. However, an issue arises in the process of local spectral convolution with high dimensional satellite image data such as hyperspectral data or multisensor data e.g. PolSAR-MS data. While performing local spectral convolution, various combinations of spectral bands are not convoluted together and therefore, under-utilizing the full potential of the spectral information present in the data. For example, it is evident from Figure 5.4(c), that the second spectral band and the tenth spectral band cannot not get convoluted together using 1D-CNN-v2 and hence the spectral information is under-utilized. The issue needs more attention for multisensor data processing with CNNs. For PolSAR-MS data, some polarimetric channels might not convolute together with some multispectral bands. For example, out of the 13 bands of the PALSAR-2-Landsat-8 dataset considered in the present study, only 4 consecutive bands get convoluted together with a convolutional kernel window of 4. Moreover, stacking PolSAR and MS data in a common feature set induces an arbitrary ordering and irrelevant band neighbourhood relationship which might impact the performance of the local spectral convolution strategy employed in 1D-CNN-v2. Different band ordering might result in different classification results. MS neighbour bands may be related by some wavelength ordering relation, but the boundary between PolSAR and MS bands is irrelevant. Therefore, it is important to perform convolutions in a manner which compensates the ordering of PolSAR-MS bands.

To provide a solution which compensates for the under-utilization of the spectral information due to PolSAR-MS band ordering, the present study proposes an ensemble approach which attempts to improves the utilization of spectral information. To present a proof of concept, this study introduces a permuted local spectral CNN model namely Perm-LS-CNN which is based on the 1D-CNN-v2 discussed in section 5.4.1. Perm-LS-CNN performs local spectral convolutions on permuted sets of the PolSAR-MS data. The permuted sets of the PolSAR-MS

data provide increased combinations of PolSAR-MS bands that are missing from using only the original set. The idea of local spectral convolution with permuted PolSAR-MS bands can be realized in Figure 5.4(d). For example, first filter (in yellow) convolves with bands 1, 3, and 7 whereas second filter (in blue) convolves with bands 2, 5, and 6. The proposed Perm-LS-CNN model configurations are provided in Table 5.3. A permutation layer before the first convolution layer is added to provide permuted sets of the original PolSAR-MS data. Although the optimal number of permuted sets required for full utilization of the spectral information is not derived in this study, 20 permuted sets are used as a proof of concept. Other configuration and hyperparameters are similar to 1D-CNN-v2 and are listed in Table 5.3.

Further, based on the Perm-LS-CNN, another novel classification framework for land cover classification using PolSAR-MS data is introduced here. This framework extends the idea of Perm-LS-CNN to spatial context. The proposed permuted local spectral-spatial CNN or the Perm-LSS-CNN utilizes the spatial feature extraction capability of CNNs, along with the idea of permuted local spectral convolution. The added spatial context helps CNNs to extract more complex and abstract features, which improves classification results. The idea of permuted spectral-spatial convolution is presented in Figure 5.5(d). For example, a first filter (in yellow) convolves with bands 1, 2, and 4 whereas a second filter (in blue) convolves with bands 5, 7, and 10. The proposed Perm-LSS-CNN model configurations are provided in Table 5.3. Similar to Perm-LS-CNN, 20 permuted sets are created in the Perm-LSS-CNN model. All hyperparameters are set as for 3D-CNN-v1 and are listed in Table 5.3.

The advantage and significance of the permuted local spectral-spatial convolutions used in Perm-LSS-CNN over the 2D-CNN-v1 and 3D-CNN-v1 is evaluated based on a class separability measure [175]. The separability measure or separability index (SI) is computed using equation (3.4). The class separabilities are measured on the ground truth class samples of Roorkee study area (refer Table 5.1) and at the first convolutional layer i.e. C1, of the 2D-CNN-v1, 3D-CNN-v1 and the proposed Perm-LSS-CNN models. For each model and class pair, the highest SI value achieved using equation (3.4) is given in Table 5.4. For example, in the BS-water class pair case, the highest SI value of; 1.82 for 2D-CNN-v1, 1.99 for 3D-CNN-v1, and 2.87 for the Perm-LSS-CNN are obtained. In other example, in the tall vegetation (denoted as TV), short vegetation (denoted as SV) class pair case, the highest SI value of; 1.46 for 2D-CNN-v1, 1.56 for 3D-CNN-v1, and 2.77 for the Perm-LSS-CNN are obtained. Overall, it is observed from Table 5.4 that the Perm-LSS-CNN model provides a feature space where the classes are more significantly separable than what is achieved by the 2D-CNN-v1 and the 3D-CNN-v1 models. The experiment establishes the advantage of permuted local convolution in the spectral dimension along with the spatial context.

Table 5.3 Perm-LS-CNN and Perm-LSS-CNN model configurations, hyperparameters and training settings.

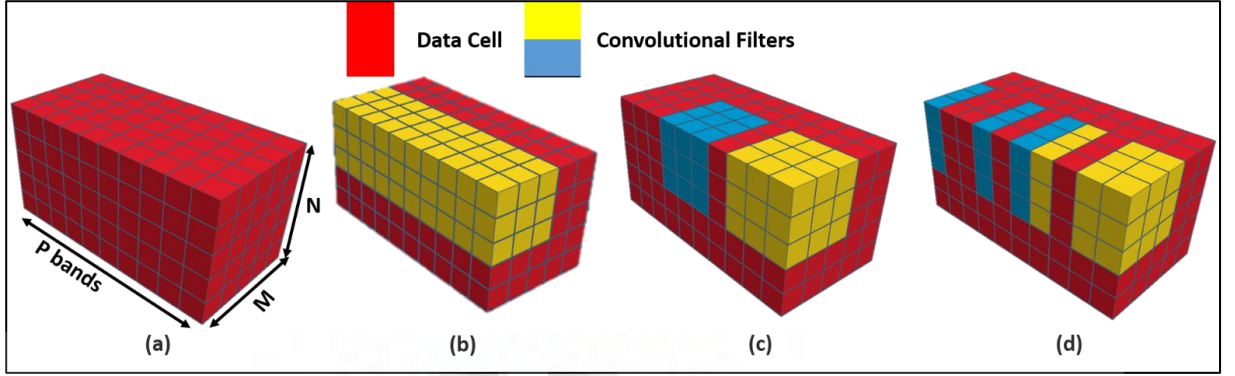| Model	Layer					
	Input	Permutation layer (P1)	Convolutional layer 1 (C1)	Convolutional layer 2 (C2)	Fully connected layer (FC)	Output layer
Perm-LS-CNN	Input image size = 1×13 Normalization = Batch Normalization	Permuted sets = 20 Kernel size = 1×13 Activation = Linear	Filters = 20 Kernel size = 1×4 Activation = ReLU Dropout fraction = 0.2 Pooling size = 0	Filters = 20 Kernel size = 1×4 Activation = ReLU Dropout fraction = 0.1 Pooling size = 0	Nodes =16 Activation = ReLU Dropout fraction = 0.1	Labels = 5 Activation = Softmax
Perm-LSS-CNN	Input image size = $3 \times 3 \times 13$ Normalization = Batch Normalization	Permuted sets = 20 Kernel size = $1 \times 1 \times 13$ Activation = Linear	Filters = 20 Kernel size = $3 \times 3 \times 4$ Activation = ReLU Dropout fraction = 0.2 Pooling size = 0	Filters = 20 Kernel size = $1 \times 1 \times 4$ Activation = ReLU Dropout fraction = 0.1 Pooling size = 0		
Training parameters	Loss function = Categorical cross entropy Learning rate = 0.01 Number of epochs = 200 Optimizer = Adam [211] 10 fold cross validation strategy is employed during training					

116

Table 5.4 Separability index (SI) values for all the possible class pairs for the 2D CNN-v1, 3D CNN-v1, and 'Perm-LSS-CNN filters.

Class pairs	Highest Separability Index (SI) value		
	2D CNN-v1	3D-CNN-v1	Perm-LSS-CNN
BS vs Water	1.82	1.99	2.87
BS vs Built-up	2.09	2.35	2.90
BS vs TV	2.98	3.82	3.75
BS vs SV	1.29	1.69	1.87
Water vs Built-up	3.05	3.04	3.57
Water vs TV	2.59	3.75	4.38
Water vs SV	1.43	1.29	1.64
Built-up vs TV	2.81	3.64	5.47
Built-up vs SV	1.70	2.10	2.27
TV vs SV	1.47	1.56	1.77

5.6. Results and Discussions

A. *Quantitative analysis*

The models 1D-CNN-v1, 1D-CNN-v2, 2D-CNN-v1, 3D-CNN-v1, Perm-LS-CNN, and Perm-LSS-CNN are trained and cross validated on ground truth class data points provided in Table 5.1. Classification results are shown in Figure 5.6 and Table 5.5 for the Roorkee data set, and in Figure 5.7 and Table 5.6 for the Haridwar data set. Figure 5.6(a) and Figure 5.7(a) show Google Earth images for the Roorkee and the Haridwar data sets respectively for reference purpose. Authors in [212] reports a detailed comparison of deep neural networks with SVMs for land cover classification with satellite image data. They deplore the lack of case studies where a quantitative and qualitative comparison of deep neural networks with SVMs with multi or high spectral satellite image data. Therefore, two non-CNN classifiers namely radial basis function-support vector machine (RBF-SVM), and random forest (RF) [33] are also considered for comparisons. The RBF-SVM classifier parameters settings are; (a). Regularization parameter is set to 12.7, and (b). Kernel influence parameter is set to 0.056568. The random forest classifier parameters settings are; (a) The number of features to consider when looking for the best split is set to 4, and (b) The number of trees in the forest is set to 41. From Table 5.5 and Table 5.6, it is evident that the proposed Perm-LSS-CNN provides the best performance for both the datasets with a 10-fold cross validation overall accuracy of 97.8 and quadratic kappa of 0.972 for the Roorkee data set, overall accuracy of 97.9 and quadratic kappa of 0.974 for the Haridwar data set. Individual class precision (a.k.a. user) and recall (a.k.a. producer) accuracies are also provided in Table 5.5 and Table 5.6 for the Roorkee and the Haridwar data sets. The difference between precision and recall averaged over all land cover classes (termed as AD in Table 5.5 and Table 5.6 is computed to evaluate the generalization ability and consistency of the classifiers. AD is minimum with Perm-LSS-CNN for Roorkee data and second minimum for the Haridwar

data. It is also observed that the AD is consistent for both data sets, i.e. around 3 whereas the AD for other classifiers fluctuates with data sets. This indicates the good generalization ability of the proposed classifier. Further, Table 5.5 and Table 5.6 also show that even if the overall accuracy of the proposed Perm-LSS-CNN is best, the other classifiers have shown good overall accuracies as well.

B. *Qualitative analysis*

A detailed analysis of the classification performance of all the classifiers considered here is however required to support these tabular results. A visual analysis of the classified images from RF, RBF-SVM, 2D-CNN-v1, 3D-CNN-v1, and Perm-LSS-CNN is performed. Figure 5.8 shows the classification results on selected areas for the five models. The selected areas are marked as A, B, C, D, and E in Figure 5.6(a). They contain the points that are taken as ground truth.

In Figure 5.8, Figure 5.8(a) show a river and man-made riverbed markers. In Figure 5.8(b, c, d, e and f), the man-made riverbed structures are correctly classified by each method. But, in Figure 5.8(b, and c), water flow in the river is not consistent which is undesirable, whereas, in Figure 5.8(d, e and f), water flow is consistent. Figure 5.8(g) shows an agriculture field with a building located in the upper right corner. In Figure 5.8(h, and i), the building is classified as SV by RF and RBF-SVM, whereas it is correctly classified as built-up by 2D-CNN-v1, 3D-CNN-v1, and the Perm-LSS-CNN in Figure 5.8(j, k and l). Figure 5.8(m) shows a dried up river, a canal, and a bridge over the canal. The river is partially classified as SV in Figure 5.8(n, and o) whereas, it is correctly classified as BS in Figure 5.8(p, q and r). The bridge is well off the ground and oriented perpendicular to the orbit of the SAR sensor, hence exhibits double bounce scattering (only from one side). It is consistently classified as a built-up structure in Figure 5.8(p and r). The same is wrongly classified in Figure 5.8(n, o and q). Observing the above two examples suggest that RF and RBF-SVM methods are unable to capture subtle changes in land covers. Also, 3D-CNN-v1 classified the bridge as Bare Soil, missing the double bounce nature inherited in the PolSAR features. Figure 5.8(s) show a peculiar TV pattern within a golf ground. The structural and spectral identity are better preserved by Perm-LSS-CNN (refer Figure 5.8(x)) as compared to other classifiers. In Figure 5.8(t, and u), most of the ground areas (bare soil or short vegetation) is wrongly classified as water. In Figure 5.8(v), the golf ground is approximately well classified, although the shape is a partially deteriorated. Figure 5.8(y) shows a tree line. Again, the structural and spectral identity is most significantly preserved by Perm-LSS-CNN as shown in in Figure 5.8(d1)).

Overall observation of Figure 5.8 suggest that the proposed Perm-LSS-CNN method yields better classification performance both in terms of classification accuracy and

generalization ability by capturing the structural and spectral identity of land covers. In fact, classification accuracies alone cannot be used as reliable indicator for comparing classification methods because of overfitting. For example, RF and RBF-SVM provides good classification accuracies (refer Table 5.5 and Table 5.6) but lacks generalization performance (refer Figure 5.8). The generalization accuracy on well-known ground truth is the real test. The reason for the inferior generalization ability of the 2D-CNN-v1 is that the 2D convolutional filter is convolving with all the spectral features, which reduces the possibility for generation of local spectrally significant features (refer Figure 5.5(b)). The proposed Perm-LSS-CNN generates such combinations of spectral and spatial properties and hence shows improved generalization performance.

In summary, the analysis of the results in Table 5.5, Table 5.6, and Figure 5.8 indicate the improved classification performance of the proposed Perm-LSS-CNN over other classifiers considered for comparison. The discussion also indicates an increase in the generalization ability of the Perm-LSS-CNN because of the ensemble learning scenario.
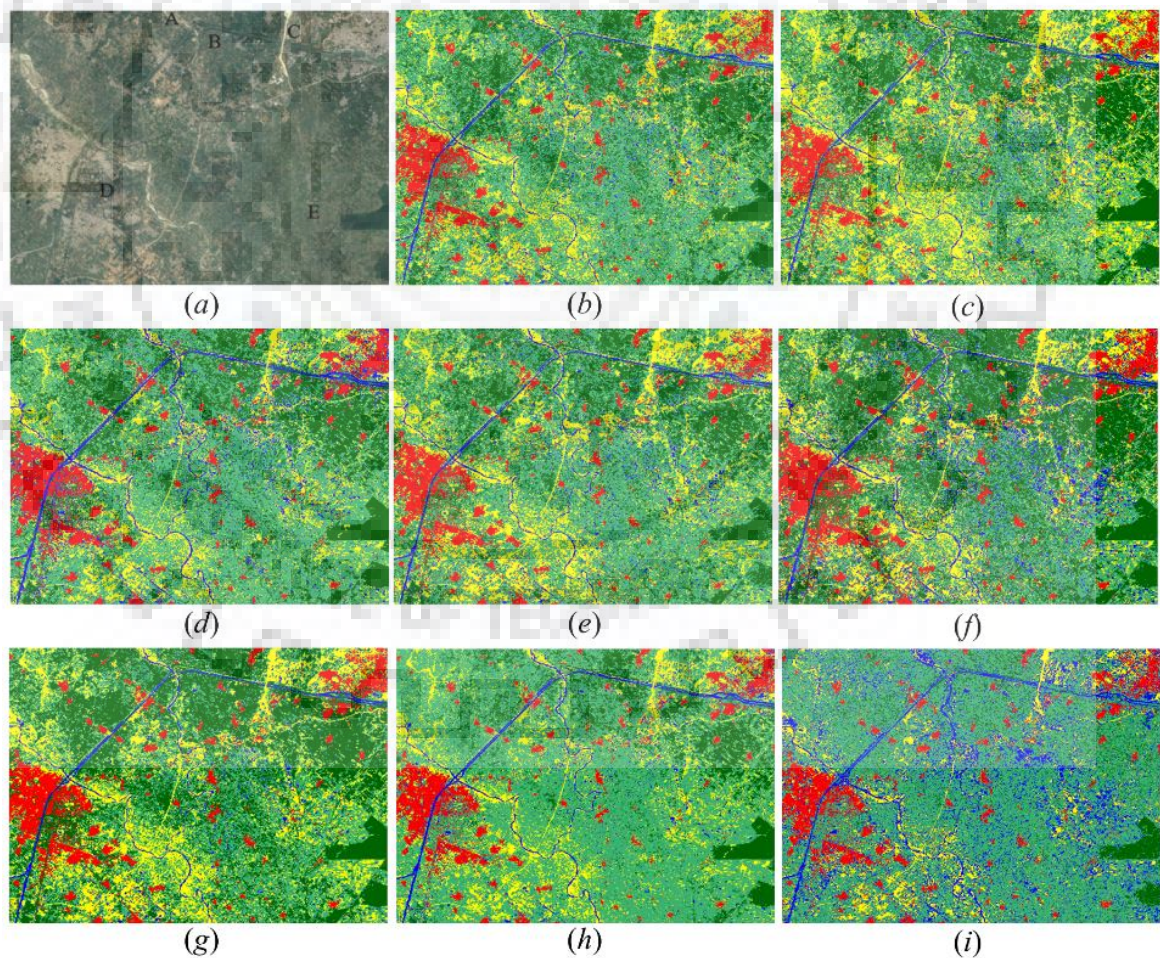
Figure 5.6 Classification image: (a) Google Earth image of the Roorkee study area, India. Classified images are shown for different classifiers: (b) 1D-CNN-v1, (c) 1D-CNN-v2, (d) Perm-LS-CNN, (e) 2D-CNN-v1, (f) 3D-CNN-v1, (g) Perm-LSS-CNN, (h) RBF-SVM, and (i) RF. Colour scheme; red = built-up, light green = short vegetation, green = tall vegetation, blue = Water, and yellow = bare soil.
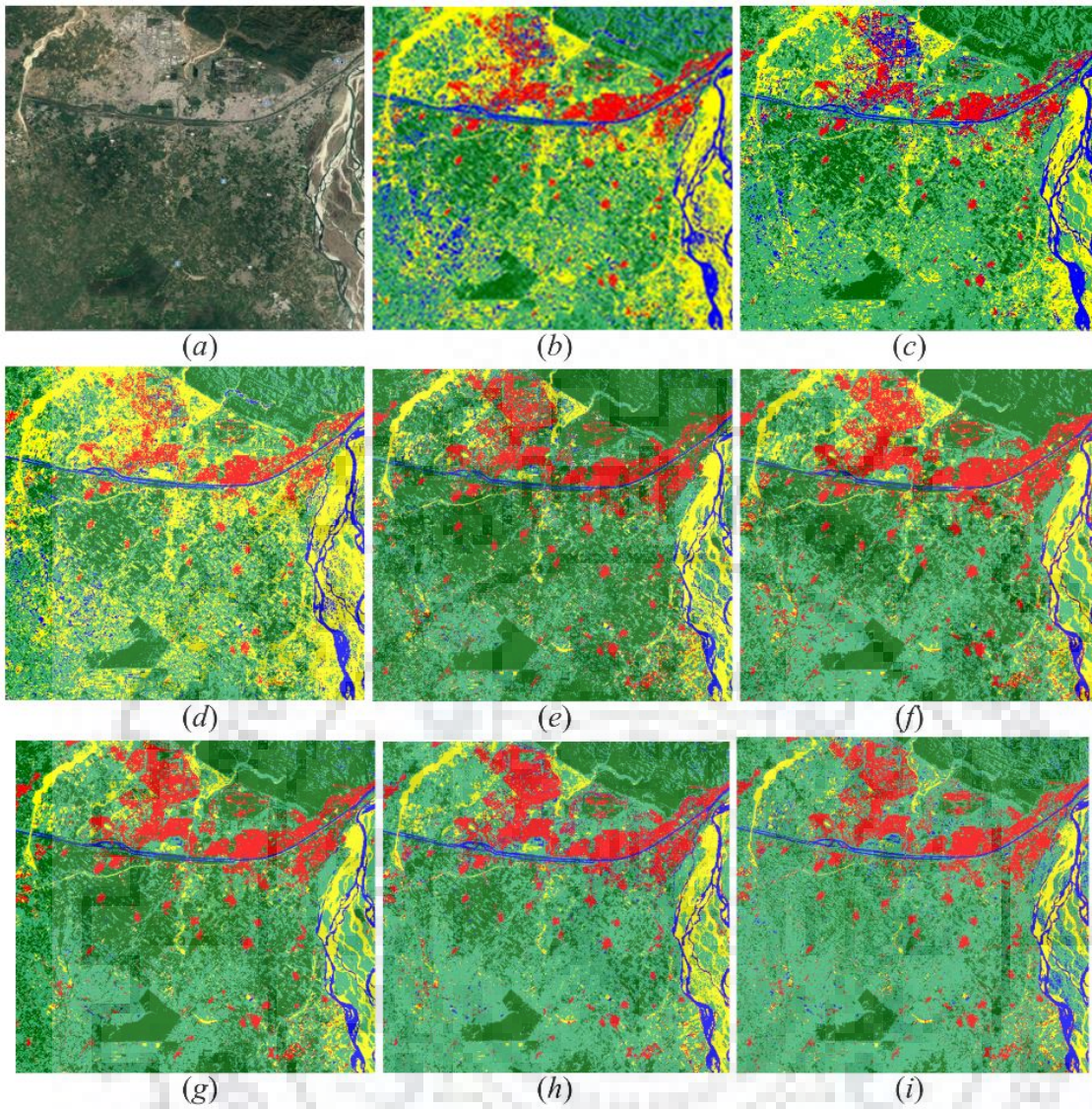
Figure 5.7 Classification image: (a) Google Earth image of the Haridwar study area, India. Classified images are shown for different classifiers: (b) 1D-CNN-v1, (c) 1D-CNN-v2, (d) Perm-LS-CNN, (e) 2D-CNN-v1, (f) 3D-CNN-v1, (g) Perm-LSS-CNN, (h) RBF-SVM, and (i) RF. Colour scheme; red = built-up, light green = short vegetation, green = tall vegetation, blue = water, and yellow = bare soil.

120

Table 5.5 Classification performance of various classifiers considered for PolSAR-MS Roorkee dataset. *AD is the difference between precision and recall, averaged over all the land cover classes considered. BS = bare soil, TV = tall vegetation, SV = short vegetation, BU = built-up, W = water.

Method	OA (%)	Kappa	User accuracy (%)					Producer accuracy (%)					AD*
			BS	TV	SV	BU	W	BS	TV	SV	BU	W	
1D-CNN-v1	95	0.938	94.5	100	85.2	100	97.5	93.4	92.4	97	97.5	96.3	4.9
1D-CNN-v2	95.2	0.939	96.4	100	84.5	100	97.1	93	91.8	96.6	98.1	98.4	5.3
Perm-LS-CNN	95.7	0.946	97	100	86.5	100	97.1	94.4	92.1	94.2	98.1	98.4	4.3
2D-CNN-v1	95.3	0.941	96.4	100	83.9	100	98.4	94.1	92.7	98	96.9	96.3	5.8
3D-CNN-v1	96.3	0.933	98.8	100	84.5	100	99.6	95.6	91	100	98.7	98.4	6.1
Perm-LSS-CNN	97.8	0.972	99.7	99.6	91.1	100	99.3	94.3	98.3	99.6	98.7	99	3.3
RBF-SVM	94.6	0.93	94.6	99.3	90.2	96.2	94	92.6	94	93.5	98	96.7	3.2
RF	91.7	0.932	86.8	96.8	88.2	96.8	93.1	92.1	97.1	65	88.6	90.6	7.9

Table 5.6 Classification performance of various classifiers considered for PolSAR-MS Haridwar dataset. *AD is the difference between precision and recall, averaged over all the land cover classes considered. BS = bare soil, TV = tall vegetation, SV = short vegetation, BU = built-up, W = water.

Method	OA (%)	Kappa	User accuracy (%)					Producer accuracy (%)					AD*
			BS	TV	SV	BU	W	BS	TV	SV	BU	W	
1D-CNN-v1	94.1	0.925	100	100	70.2	96.6	99.3	97.1	87.1	98.9	95	99.3	9.2
1D-CNN-v2	95	0.937	100	100	73.2	99.1	99.3	99.2	87.8	100	95.9	99.3	8.6
Perm-LS-CNN	96.1	0.95	100	100	77.8	100	100	100	89.2	100	97.5	100	7.1
2D-CNN-v1	94.7	0.933	100	100	70.9	99.6	100	97.8	89.2	100	92.9	99.3	9.9
3D-CNN-v1	96.5	0.955	100	100	80.1	99.1	100	98.5	90.3	99	100	100	6.2
Perm-LSS-CNN	97.9	0.974	100	100	89.3	99.1	100	99.2	96	99.1	96.7	100	3.4
RBF-SVM	96.1	0.95	98.5	100	86.2	95.8	97.2	97.8	92.7	95.7	96.6	100	4.3
RF	94.4	0.928	97	98.6	86.2	91.6	95.1	97.7	95.9	84.9	91.6	100	1.9
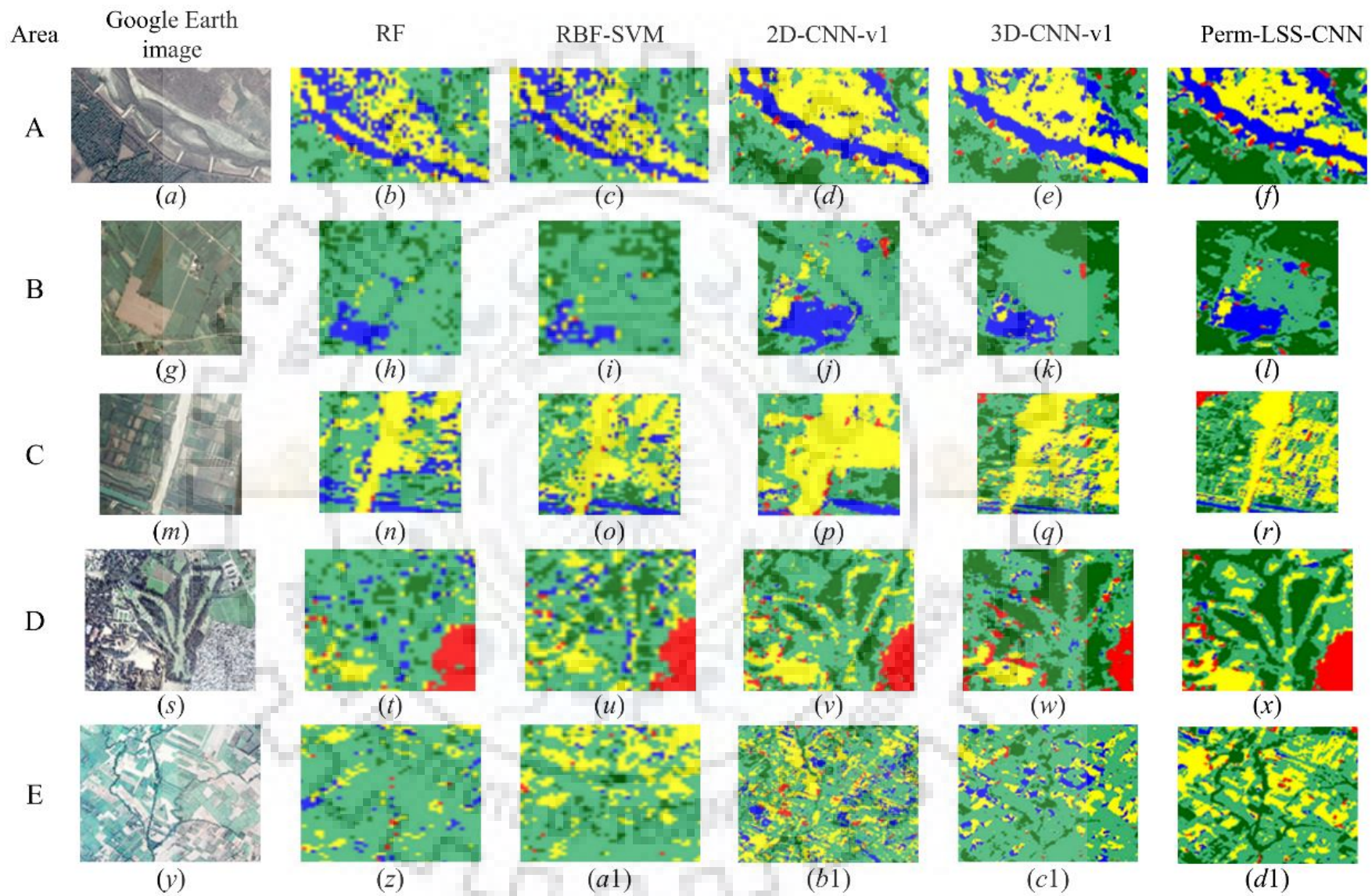
Figure 5.8 Results from different classification techniques applied on the Roorkee data set. The areas include data points collected as ground truth.

5.7. Conclusion

This chapter presents a critical analysis of the potential and performance of the one, two, and three dimensional CNNs, where dimensionality refers to the convolved features: spectral, spatial or both. These CNNs are applied to land cover classification with multisensor (PolSAR-MS) satellite image data. Two novel models are introduced, where local filters in either spectral domain (Perm-LS-CNN) or spatial-spectral domains (Perm-LSS-CNN) are applied after a permutation layer. The permutation layer generates an ensemble of band combinations, which are drawn upon by the CNN in order to generate powerful features for land cover classification. Classification results demonstrate the improved generalization ability of these models compared to both alternative CNN architectures and other machine learning models (SVM and random forests). Perm-LSS-CNN also fully uses both PolSAR and multispectral information, improving results compared to other popular methods for PolSAR-MS data based land cover classification. A separability measure is used for the model significance evaluation. Separability analysis states that the Perm-LSS-CNN transforms the input data space into a feature space where separability among classes increases. This increase in separability is achieved due to the inclusion of permuted local spectral convolution along with the local spatial convolution which is not available in the 1D-CNN-v1, 2D-CNN-v1, or 3D-CNN-v1 CNN models. A fine analysis on ground truth data provides an extra qualitative assessment which complements the quantitative classification results. These qualitative assessments demonstrate that the proposed Perm-LSS-CNN model is able to generalize much better while keeping high classification accuracies.

CHAPTER 6

NOVEL CNN, RNN, AND CNN-RNN MODELS FOR MULTISENSOR SATELLITE TIME SERIES IMAGE BASED CROP CLASSIFICATION

Crop classification is an important task in many crop monitoring applications. Satellite image based monitoring provides an easy, techno-economical, reliable, stable and fast approaches to crop monitoring tasks. Numerous techniques have been reported for satellite time series image based crop classification. However, with the availability of more and more high quality satellite images from new and powerful sensors, which are increasing by the day, it has become a challenge for techniques to keep up in terms of performance, high dimensional operability, scalability, and generalizability. The new technologies must should have a unified framework to be able to process plethora of multi-resolution, multisensor, and high dimensional satellite time series images. In the context, CNNs and recurrent neural networks (RNNs) are the most recent and trending deep neural network frameworks for satellite time series image based agriculture crop monitoring applications. Both, CNN and RNN models have been developed in, standalone and synergistic modes for agriculture crop classification by many researchers. However, the CNN and RNN models are still young, promising, and growing techniques in the satellite image proceeding community especially in multisensor satellite time series image processing and therefore further exploration of the potential of these models is motivating. Also, certain limitations such as presence of mixed land cover scenario still exist. Mixed land cover scenario in crop classification refers to the presence of other land covers (built-up, forests, or grasslands) in-between and around the croplands in the study region (e.g. in India). Therefore, it is still a challenge for existing DNN based techniques to synergistically exploit the spatial, spectral, and temporal information of a crop present in the multimodal and/or multisensor satellite time series images and provide accurate and reliable crop type information in mixed land cover scenarios.

This chapter presents a study in which an attempt is made; to analyse conventional, and to develop novel, CNN and RNN models based agriculture crop classification models with multisensor/multimodal satellite time series imagery. This chapter is an extension of chapter 5 in a sense that the concept of permuted spectral input is used here with time series images and how DNN models exploit the satellite images with spectral, spatial, and temporal information.

6.1. Introduction

Recent reports indicate the initiatives taken by authorities worldwide towards development of sustainable agriculture science and management strategies and policies [213]–[215]. Satellite image based crop monitoring is a key technology in acquiring crop information such as crop type, or crop phenology[14], [216]–[218]. Satellite image based crop monitoring is easy, fast, reliable, techno-economical, and efficient [219]–[221]. In the recent years, India is strongly leaning towards the use of satellite images for crop monitoring and crop information extraction [222], [223] but still needs to go far.

Optical and PolSAR satellite images have been successfully used for agricultural applications in India such as crop mapping/classification [224], crop yield forecasting [225], crop stress and drought stress monitoring [226], mapping of crop type/cycle changes [227], crop phenology assessment [228], and biomass estimation [229] in the last two to three decades. The use of satellite images in these applications is increasing day by day and new applications such as precision agriculture [230] are being discovered as well. In particular, with the availability of open source high spatial and/or spectral and/or temporal resolution multisensor satellite images via programs such as the Copernicus Sentinel satellite series, the performance and efficiency of the aforementioned applications has increased. Also, new and advanced studies in crop monitoring have surfaced [72], [231], [232]. Recent studies on Indian croplands are also using Sentinel series data for crop monitoring [233], [234].

Moreover, time series satellite images are rich source of crop phenology information and have become popular, efficient, and crucial for agriculture crop monitoring including crop identification and classification. For agriculture, it has been shown that the temporal evolution of the signal measured via satellite imaging enables a better separation between crop types classes [235]. Satellite time series image based agriculture crop classification is one of the most cost-effective, reliable, and successful framework. Also, different image modalities have been successfully used in crop classification. Satellite PolSAR imaging is a weather independent technology and time series PolSAR images have been successfully used in crop classification applications[236]–[240]. On the other hand, time series optical imagery provides crop phonology information via spectral-temporal features. Time series optical imagery has also been successfully used for crop classification in the recent past [241], [242] The reduced number of optical data based crop classification studies is because of the effect of clouds on optical satellite images. PolSAR and optical image modalities have also been used synergistically for crop classification to improve the performance of methodologies [53], [67], [243]–[248].

In context to satellite time series image processing techniques for crop classification, numerous techniques have been proposed and/or utilized in the past. The popular techniques are maximum likelihood classifier [238], [239], [243], [245], neural networks [239], [243], [248], decision trees [239], [243], random forest [67], [238], [241], [242], [246], [247], temporal fusion classification [53], and hierarchical classification strategy [244], It is observed that random forest, maximum likelihood classifier, and neural networks are the most popular methods utilized in satellite time series image based crop classification applications. However, with the availability of more and more high quality satellite images from new and powerful sensors, which are increasing by the day, it has become a challenge for these techniques to keep up in terms of performance, high dimensional operability, scalability, and generalizability. Therefore, it is important to level up the technologies. The technologies must be able to process plethora of multi-resolution, multisensor, and high dimensional satellite time series images.

In this regard, recently, deep learning with deep neural networks is one of the most promising techniques. Numerous successful approaches have been reported in the recent years which are based on deep neural networks for satellite time series image applications. In particular, CNNs and recurrent neural networks (RNNs) are the most popular deep neural network types utilized in satellite time series image based agriculture crop monitoring applications. Further, novel CNN and RNN models have also been developed and utilized in both, standalone and synergistic, modes for agriculture crop classification by many researchers. A detailed review on popular CNN and RNN models for agriculture crop monitoring and classification applications is presented in chapter 2 (refer section 2.3). The review indicated to the success and significance of deep neural network models in satellite time series image based crop classification applications. However, the technology is still young and growing, and there is scope for much more. In particular, the DNN model based studies for time series satellite image processing discussed in section 2.3 are focused more on either the spectral, spatial, or temporal characteristic or at most more focussed on two characteristics at a time ignoring the third characteristic. For example, some CNN and RNN studies focused on spectral features more than temporal or spatial features during classification which lead to under-utilization of temporal and spatial features during crop classification. Similarly, some other studies focused more on temporal and spatial characteristics and less on spectral characteristic which again lead to under-utilization of spectral features of the crops. This imbalance, in priority, requires investigation on the subject and demands novel CNN and RNN model based approaches that can address this issue. To this day, only few studies focussed on utilizing all three characteristics have been reported but a trend is underway. Other major limitation during crop classification is the presence of mixed land cover scenario i.e. presence of other land covers (built-up, forests, or grasslands) in-between and around the

126

croplands in the study region (e.g. in India). Therefore, it is still a challenge for existing DNN based techniques to synergistically exploit the spatial, spectral, and temporal information of a crop present in the multimodal and/or multisensor satellite time series images and provide accurate and reliable crop classification.

The above arguments motivate towards new studies focused on using multisensor satellite time series images, especially optical and PolSAR, for crop classification and development of novel CNN and RNN models which focus on the judicious exploitation of the spectral, temporal and spatial characteristics of crops to achieve improves classification results.

In the study presented in this chapter, an attempt is made; to analyse the conventional CNN and RNN models, and to develop novel CNN and RNN (in standalone as well as synergistic modes) models for agriculture crop classification with multisensor/multimodal satellite time series (MSTS) imagery. For this purpose, two sub tasks are set in this study, –*first,* critical analysis of conventional CNN and RNN models for MSTS images based crop classification, and –*second,* development of novel CNN and RNN models for MSTS images based crop classification. In the first sub task i.e. *sub task 1*, conventional CNN and RNN models are analysed and evaluated for the task of MSTS images based crop classification. 1D, 2D, and 3D CNNs, RNNs, and combinations of these are considered for analysis. Then, based on the critical analysis of the working and performance of these conventional models over MSTS images based crop classification, few novel models are proposed and developed in *sub task 2*.

The novel models involve permuted spectral and/or spatial convolutions and/or bi-directional recurrent layers. The Bidirectional RNNs capture the temporal profile and the permuted spectral and/or spatial CNNs capture the spectral and/or spatial profile of a particular crop. Bidirectional strategy in RNNs enables the influence of the past, present, and future inputs on the current output. Bidirectional RNNs are beneficial in time series images based crop classification because crop classification can utilize the bidirectional process i.e. classification result can be observed from any time node. Also, the permuted spectral band stacking strategy developed in chapter 5 is also used in this study and presumed to improve the significance of the spectral information present in multisensor satellite time series images. In particular, a synergistic utilization of permuted spectral-spatial CNNs and Bidirectional RNNs helps in better exploitation of the multisensor satellite time series images.

This chapter is organized as follows. Section 6.2 provides conceptual background on CNNs and RNNs. Further, section 6.3 provides the information on the experiment setup i.e. study areas, datasets, and ground truth. Section 6.5 and 6.7 provides the implementation details on the

conventional and novel CNN-RNN models Section 6.8 discusses on results obtained using these models and its analysis. Finally, section 6.9 concludes the study.

6.2. Theoretical Background

Basic knowledge of CNNs and RNNs is required in this chapter as CNNs can exploit the spectral and/or spatial information and RNNs can exploit the temporal information, present in the MSTS images. The brief concepts of CNNs are presented in chapter 3 (refer section 3.2.2) and chapter 5 (refer section 5.2.1). In addition, the basic concepts on RNNs are discussed in this section.

6.2.1. Recurrent neural networks

Recurrent neural networks, also known as RNNs, are a class of neural networks that allow previous outputs to be used as inputs while having hidden states. The RNNs have the ability to process input of any length while keeping the model size unchanged. RNNs also share weights across time along with space. Therefore, the output depends on historical information. A typical architecture of a RNN is shown in Figure 6.1.

For each timestep t, the activation $a^{<t>}$ and the output $y^{<t>}$ are expressed as follows [249].

$$a^{<t>} = g_1\left(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a\right) \tag{6.1}$$

$$y^{<t>} = g_2(W_{ya}a^{<t>} + b_y) \tag{6.2}$$

where W_{aa}, W_{ax}, W_{ya} b_a, and b_y are coefficients that are shared temporally, g_1 is an activation function (usually a hyperbolic tangent), and g_2 is a softmax activation function [249].
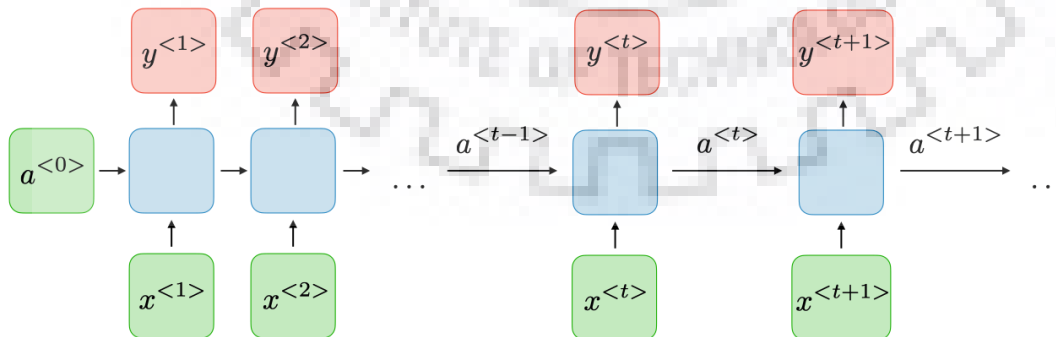
Figure 6.1 Typical architecture of a recurrent neural network [250].
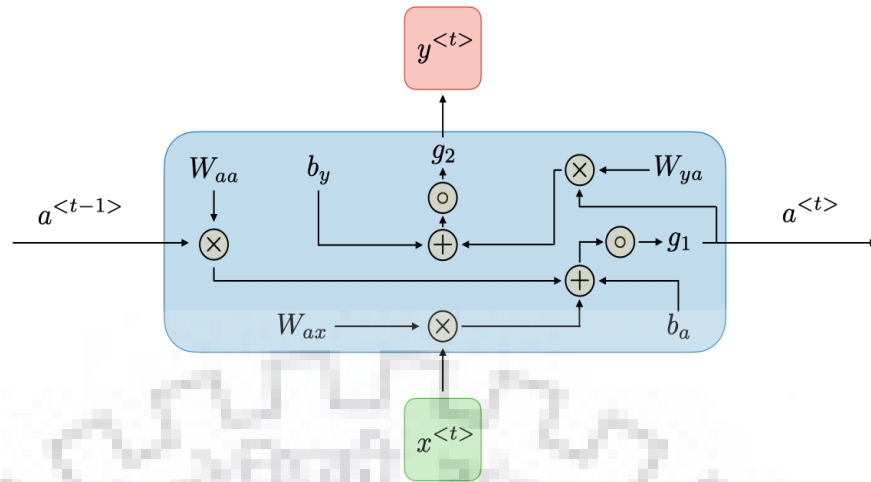
Figure 6.2 Signal flow graph in a simple recurrent cell [250].

A. Recurrent cell

A recurrent cell is the smallest unit in RNNs which is able to memorize activation representing historical information. There are different variants of recurrent cells. The popular recurrent cell architectures are discussed here.

Simple recurrent cell

Architecture and signal flow of a simple recurrent cell is shown in Figure 6.2. The cell takes $x^{<t>}$ i.e. current input and $a^{<t-1>}$ i.e. activation from previous cell, as inputs and outputs $a^{<t>}$ i.e. activation for next RNN cell and $y^{<t>}$ i.e. prediction from current state. Although the cell is able to capture short historical information but is unable to retain longer historical information. There are two popular variants of the basic RNN cell that are able to capture longer historical relations within a signal.

Long short term memory (LSTM) recurrent cell

Recurrent neural networks with simple recurrent cell have problems associated with vanishing and exploding gradients [251]. This is a common problem in neural network updates where successive multiplication by the weight matrix W is inherently unstable; it either results in the gradient disappearing during backpropagation, or in it blowing up to large values in an unstable way. In RNNs, this type of instability is the direct result of successive multiplication with the (recurrent) weight matrix at various time-stamps. One way of viewing this problem is that a neural network that uses only multiplicative updates is good only at learning over short sequences, and is therefore inherently endowed with good short-term memory but poor long-term memory [251]. This problem can be addressed with the use of a long short-term memory

129

(LSTM) cell for longer memory. The operations inside the LSTM cell are designed to have fine-grained control over the data written into this long-term memory. Figure 6.3 show the internal architecture and signal flow in a LSTM recurrent cell. The LSTM recurrent cell has a new variable $c^{<t>}$, which stands for memory cell at time t. The memory cell will provide memory to remember activations which can be used during predictions at later time steps.

The governing equations inside a LSTM recurrent cell are as follows [252].

$$\tilde{c}^{<t>} = \tanh\left(W_c\left[\Gamma_r * a^{<t-1>}, x^{<t>}\right] + b_c\right) \tag{6.3}$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>} \tag{6.4}$$

$$a^{<t>} = \Gamma_0 * c^{<t>} \tag{6.5}$$

Where, $\tilde{c}^{<t>}$ is a candidate for being a current memory cell state which is based on, relevance gate i.e. Γ_r status, previous cell activation i.e. $a^{<t-1>}$, and current input i.e. $x^{<t>}$. W_c and b_c are trainable parameters. $c^{<t>}$ is the current memory cell state which is based on, update gate i.e. Γ_u status, forget gate i.e. Γ_f status, candidate for memory current cell state i.e. $\tilde{c}^{<t>}$, and old memory cell state i.e. $c^{<t-1>}$. $a^{<t>}$ is current cell activation which is based on, output gate i.e. Γ_0, and current memory cell state i.e. $c^{<t>}$ [252].

Gated recurrent unit (GRU) recurrent cell

The Gated Recurrent Unit (GRU) can be viewed as a simplification of the LSTM, which does not use explicit cell states. Another difference is that the LSTM directly controls the amount of information changed in the hidden state using separate forget and output gates. On the other hand, a GRU uses a single reset gate to achieve the same goal. The architecture and signal flow in a GRU recurrent cell is shown in Figure 6.4.

The governing equations inside a GRU cell are as follows [253].

$$\tilde{c}^{<t>} = \tanh\left(W_c\left[\Gamma_r * a^{<t-1>}, x^{<t>}\right] + b_c\right) \tag{6.6}$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) * c^{<t-1>} \tag{6.7}$$

$$a^{<t>} = c^{<t>} \tag{6.8}$$

Where, all the parameters have usual meanings. The GRU cell replaces the forget gate Γ_f with $1 - \Gamma_u$ and does not utilizes the output gate Γ_0 [253].

130
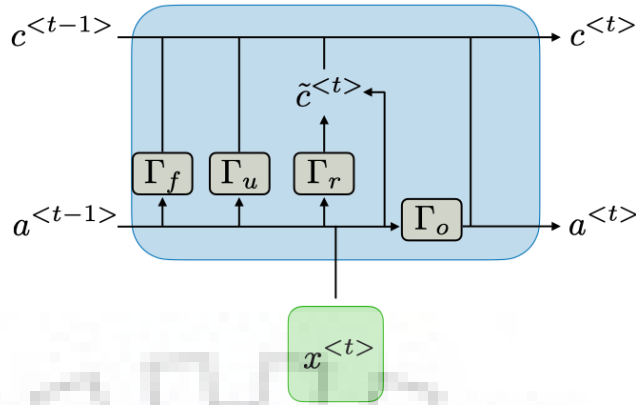
Figure 6.3 A LSTM recurrent cell [250].
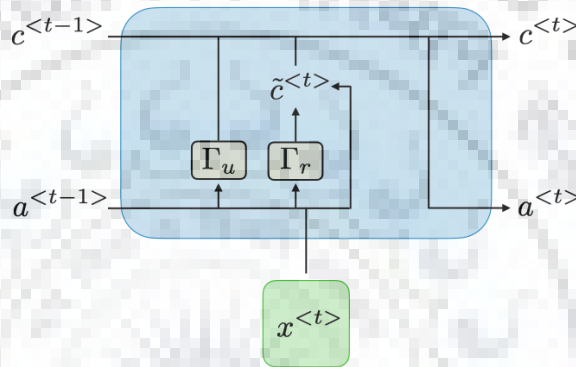
Figure 6.4 A GRU recurrent cell [250].

B. Learning in RNNs

The key concept behind learning in RNNs is the 'backpropagation through time' algorithm. In backpropagation through time, the negative logarithms of the softmax probability of the correct predictions at the various time-stamps are aggregated to create the loss function. If the softmax probability at time t is represented as $p^{<t>}$, the loss L is defined as follows [252].

$$L = -\sum_{t=1}^{T} \log\left(\hat{p}^{<t>}\right)$$

(6.9)

C. RNN architectures

RNN models are mostly used in the fields of time series signal processing. Overall, there are four variants of RNN architectures.

i.   One to one
ii.   One to many
iii. Many to one
iv.   Many to many

The many to one and many to many versions of RNNs shown in Figure 6.5 are popularly used for crop classification in many studies.
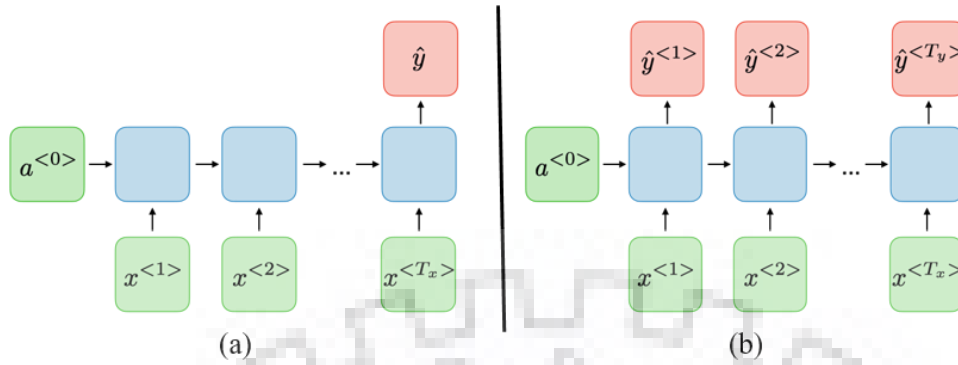
Figure 6.5 RNN architectures used popularly in satellite time series image based crop classification: (a) Many to one, and (b) Many to many RNN architectures [250].

In the current study, multiple versions of RNNs, alone and in combination with CNNs are explored. Details of the CNN and RNN models developed and utilized in the study are provided in further sections.

6.3. Experiment Setup

This section provides information about the resources collected and utilized during the study. It includes information of study areas, datasets, and ground truth selected or collected for the study.

6.3.1. Study area

The study area selected for study is a cropland region in the Haridwar district of the Uttarakhand, India where common seasonal crops are practiced. The region is considered based on accessibility to ground truth collection, crop season, and generic land cover distribution. The geographical extents of the region are provided in Table 6.1. Figure 6.6 shows the Google Earth image of the study area. The region is a land parcel which is managed by local farmers. Seasonal crops and sugarcane are grown in this region. The study area is approximately a 6km by 7km land parcel, covering $42km^2$ of area. The region also has forests, grasslands, deciduous tree plantations along with various crops in the study area. This makes the region more realistic environment and mixed class scenario for algorithm testing. The presence of multiple land covers will evaluate the generalization ability of the classifiers. Areas marked in the image (1, 2, 3, and 4) are subsets of study area used for qualitative and visual evaluation of the classification performance of classifiers developed in the study.

Table 6.1 Geographical extent of study area.

Coordinates (upper left)		Coordinates (lower right)	
Latitude	Longitude	Latitude	Longitude
29.84183	77.02115	29.78843	78.09080

Figure 6.6 Google Earth image of the study area (GE Image Copyright 2018). Subset areas selected for ground truth via visual inspection. 1: GTI-1, 2: GTI-2, 3: GTI-3, and 4: GTI-4 (highlighted in yellow boxes).

6.3.2. Dataset

Multisensor satellite images are considered for crop classification in this study. Sentinel-1 PolSAR and Sentinel 2 multispectral satellite sensor images are used. Sentinel-1 C band dual pol SAR images are available at 10 m spatial resolution and at a revisit time of 6 days (12 if pass direction is considered). Similarly, Sentinel-2 multispectral images are available at 10m, 20m, and 60m spatial resolutions and at a revisit time period of 5 days Both sensor datasets are freely available for public use. Based on the growing season of wheat, Sentinel-1 ground range detected (GRD) and Sentinel-2 L1C satellite time series images are collected every month throughout the wheat growing season for year 2017-18. Acquisition dates of each month for both sensors are listed in Table 6.2. Also,

Table 6.3 lists the specifications of bands considered from Sentinel-1 and Sentinel-2. Overall, 10 multispectral bands and 2 PolSAR channels are considered for the study. Also, 7 time steps from October 2017 to April 2018 are considered for the study. Figure 6.7 shows a false colour composite of the dataset for the study area and highlights the varied crop types.

Table 6.2 Sentinel-1 and Sentinel-2 data acquisition dates.

Month	Acquisition dates	
	Sentinel-1	Sentinel-2
October	3/10/2017	7/10/2017
November	20/11/2017	21/11/2017
December	2/12/2017	1/12/2017
February	12/2/2018	9/2/2018

133

March	8/3/2018	6/3/2018
March	1/4/2018	31/3/2018
April	13/4/2018	15/4/2018

Table 6.3 Band specifications of Sentinel-1 and Sentinel-2.

Sensor	Band selection	Spatial resolution (meters)
Sentinel-1	VH	10
	VV	
Sentinel-2	Band 2: Blue	10
	Band 3: Green	
	Band 4: Red	
	Band 8: NIR	
	Band 5: Vegetation Red Edge	20
	Band 6: Vegetation Red Edge	
	Band 7: Vegetation Red Edge	
	Band 8A: Vegetation Red Edge	
	Band 11: SWIR-1	
	Band 11: SWIR-2	
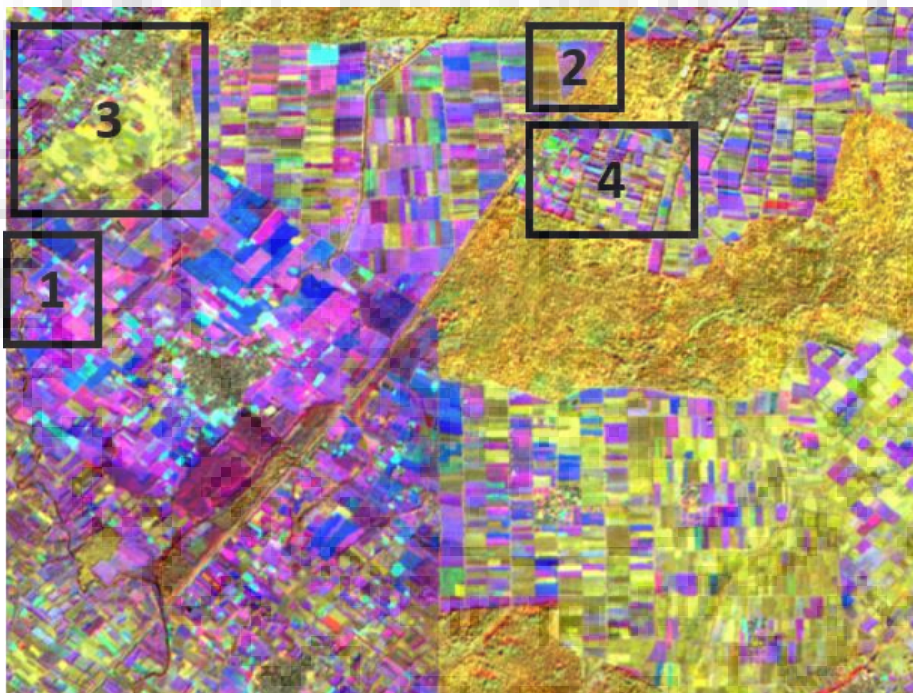
Figure 6.7 False colour composite from the Sentinel-2 time series dataset. Colour scheme; red: Band 4 (November 21, 2017), green: Band 4 (February 9, 2018), blue: Band 4 (March 31, 2018). Image subset areas selected for ground truth via visual inspection (highlighted in black boxes). 1 = GTI-1, 2 = GTI-2, 3 = GTI-3, and 4 = GTI-4.

6.3.3. Ground truth

Based on the study regions and crop season selection, wheat is the major crop grown in the area. However, the region also grows sugarcane throughout the year. The region also has evergreen forests, deciduous commercial plantations such as Salicaceae (Popular Eucalyptus), grasslands,

134

and villages which must be considered during crop classification as it is a usual landscape (mixed land cover) in India. Therefore, overall, six classes are considered namely: wheat, sugarcane, forests, deciduous plantation mostly 'Popular' plantations, grassland, and built-up. The ground truth samples are either measured directly on the terrain or are collected via visual inspection from the dataset. Summary of ground truth samples collected for all considered classes is provided in Table 6.4. Few images from the ground truth survey are also shown in Figure 6.8. Four subset areas from the study area are selected for visual and qualitative evaluation. Figure 6.6 and Figure 6.7 shows the four marked areas. The subset areas are termed as ground truth image (GTI) 1 through 4 i.e. GTI-1, GTI-2, GTI-3, and GTI-4.

Table 6.4 Summary of ground truth samples collected during ground truth survey.

Class	Ground truth samples
Wheat	300
Sugarcane	300
Forest	300
Popular	300
Grassland	300
Built-up	300

135
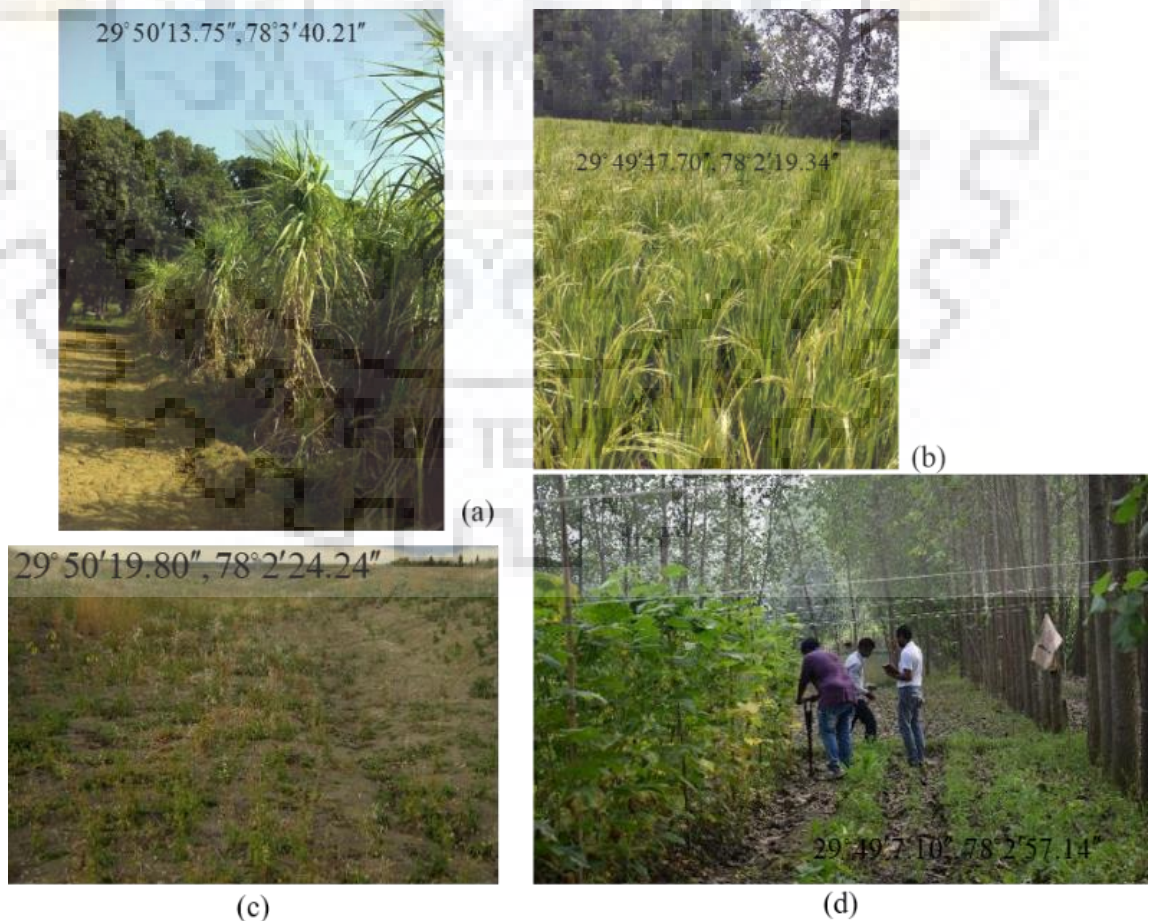
Figure 6.8 Sample images of different land covers collected during ground truth survey. (a) Sugarcane cropland, (b) Wheat cropland, (c) Grasslands, (d) Popular plantations.

6.3.4. Preprocessing

Both Sentinel-1 and Sentinel-2 datasets require preprocessing before classification. Sentinel-1 GRD datasets are terrain corrected and calibrated. Terrain correction is performed using the SRTM 1 arc-second digital elevation model (DEM) available for public use. Speckle filtering is not performed as speckle noise is already suppressed in GRD products. Sentinel-2 L1C datasets are atmospherically corrected using the 'sen2cor' processor [254]. Further the 20 m spatial resolution bands in the bottom of atmosphere (BOA) Sentinel-2 datasets are resampled to 10 m using the 'super-resolution' approach proposed in [255]. Overall, 2 polarimetric channels i.e. VH, and VV from the terrain corrected and calibrated Sentinel-1 and 10 multispectral bands from resampled bottom-of-atmosphere Sentinel-2 datasets are selected and stacked together to form a PolSAR-multispectral multisensor dataset. Overall, 7 multisensor datasets are created corresponding to the 7-time steps. Further, the 7 multisensor datasets are stacked along the temporal dimension creating a multisensor time series dataset. This PolSAR-multispectral multisensor satellite time series (MSTS) image dataset is utilized for crop classification.

6.4. Generalized DNN Framework for Satellite Time Series Images based Crop Classification

Although, numerous studies reported on satellite time series image processing have developed and utilized different DNN models, these models can be consolidated into a generalized framework. Figure 6.9 depicts a generalized DNN framework for time series satellite image based crop classification. The framework is composed of CNNs (1D, 2D or, 3D), RNNs (Simple, LSTMs or, GRUs), and fully connected (FC) modules. Different DNN architectures can be created using either any one, more than one or, all three modules. For example, 1D-CNNs can be used as spectral or temporal feature extractors with or without FC layers. Further, these 1D features can be utilized with a classifier (e.g. a softmax layer, SVM or any other) to achieve classification. Alternatively, the 1D CNNs can also be integrated with RNNs as, the former acts as spectral feature extractors and the latter acts as temporal feature extractors. The classification stage can be a softmax layer.

Similarly, 2D-CNNs can replace 1D-CNNs in the aforementioned framework. The only difference is that 2D-CNNs provide spatial features in contrast to 1D-CNNs (which provide only spectral features). If 2D-CNNs are used with RNNs then spatial-temporal features are generated

136

which can be used with a classifier to achieve classification. 1D- and 2D-CNNs can also be used in parallel and cascade fashion to obtain both spectral and spatial or, spectral-spatial features respectively. These spatial and spectral or spectral-spatial features can be used with RNNs to obtain, spectral, spatial, and temporal features, spectral-spatial-temporal features or, any combination of these. These strategies lead to full exploitation of time series images with good spatial and spectral information.

However, the cascade fashion of 1D- and 2D-CNNs can be replaced with 3D-CNNs. 3D-CNNs have the ability to extract spectral and spatial features simultaneously from satellite images. 3D-CNNs can be used with or without FC layers prior to classification. Also, 3D-CNNs can be utilized with RNNs to capture all three i.e. spectral, spatial, and temporal information from satellite time-series images. The classification stage is usual.

The 1D-CNNs can be employed at pixel level for classification whereas the 2D- and 3D-CNNs can only be employed at image patch or image window level. Per-pixel classification is not an option with 2D- and 3D-CNNs. Also, the 1D-, 2D-, and 3D-CNNs have to be employed in time distributed fashion before using RNNs as temporal feature extractors. RNNs can also be used in standalone mode to extract temporal features directly from satellite time series images at pixel level.

The CNNs, RNNs, and the FC modules can be used in numerous other fashions to achieve classification. The next few sections present few such possibilities and configurations for multisensor satellite time series image based crop classification.

Although the current study used multisensor satellite time series (MSTS) images, however, the architectures discussed above are independent of thee notion of multisensor. This means that all or any other variation of the aforementioned DNN architectures will work with multisensor satellite time series images in the similar fashion.
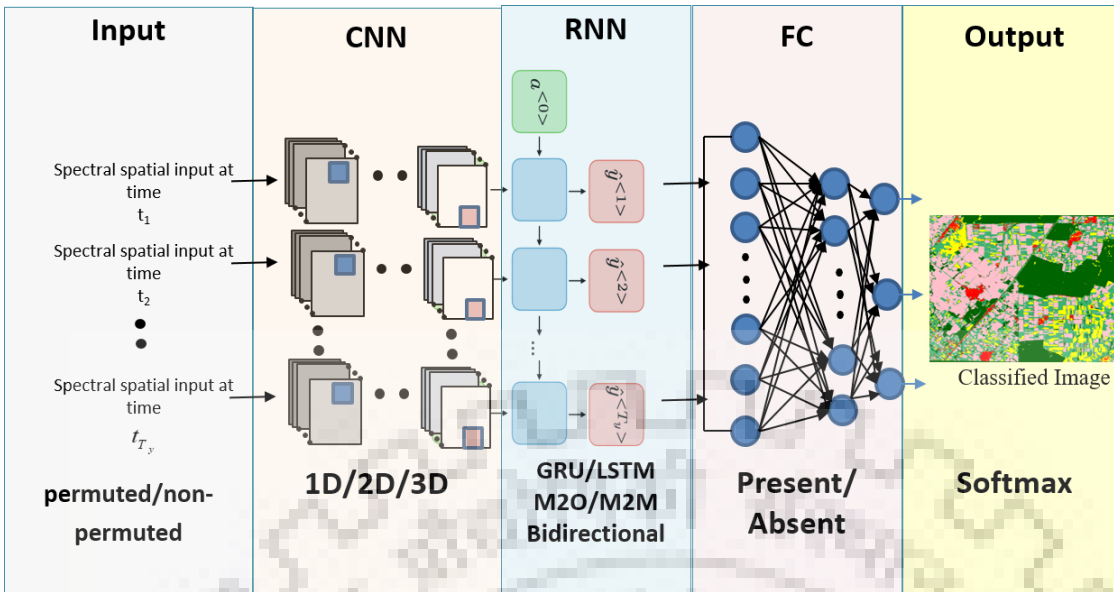
Figure 6.9 A generic framework for DNN models for satellite time series images based crop classification.

6.5. Sub Task 1: Conventional CNN and RNN models for multisensor satellite time series image based crop classification

In this section, various conventional CNN and RNN models, in standalone or synergistically, are used for multisensor satellite time series (MSTS) image based crop classification. The section provides information about the architecture, and training parameters of the models used.

6.5.1. 1D-CNN model for MSTS images based crop classification

The present study focuses on the analysis of the performance of 1D-CNNs for MSTS image based crop classification. In the 1D-CNN model namely 1D-CNN-v1 presented here, one dimensional convolution can be performed in either the spectral, or temporal dimension. Since convolution in the spectral dimension is already covered in chapter 5, here, one dimensional convolution is performed in the temporal dimension. The kernel size of the filter in the convolutional layer i.e. C1 layer is equal to the temporal depth of MSTS image i.e. 7. One dimensional temporal convolution is similar to one dimensional spectral convolution in the implementation aspect (refer Figure 5.4(b)). However, the implications and interpretations are clearly different. One dimensional local temporal convolution is not considered as the temporal depth is low.

The 1D-CNN-v1 architecture and hyper-parameter settings are listed in Table 6.5. Rectified linear unit (ReLU) and softmax activation functions are used in the intermediate and output layers respectively. A dropout strategy is considered for regularization and a dropout

fraction value of 0.2 for C1 is set. Pooling is not considered because it hinders the full usability of the temporal variation. The loss function used is the categorical cross entropy function. It optimizes the classification accuracy based on probabilities of each class during classification. The learning rate and the number of epochs are set to 0.01 and 200 respectively. It is to be noted here that the models (conventional or novel) discussed in this chapter have similar hyperparameters.

6.5.2. 2D-CNN model for MSTS images based crop classification

The input MSTS image data has 1 spectral, 2 spatial and 1 temporal dimension. It is not possible to realize 2D-CNNs, in standalone mode, to process this 4-dimensional data directly. Therefore, 2D-CNN model is not explored further.

6.5.3. 3D-CNN model for MSTS images based crop classification

The 3D-CNN model proposed here operates (window based convolution) along the time and spatial dimensions. The spectral dimension corresponds to the channels in the input data during the convolution.

In the 3D-CNN model namely 3D-CNN-v1, the size of the three dimensional kernel of the filter in the convolutional layer i.e. C1 layer is set equal to the temporal depth of MSTS image i.e. 7 in the temporal dimension and is set equal to 3×3 in the spatial dimension. Three dimensional temporal-spatial convolution is similar to spectral-spatial convolution shown in Figure 5.5(c) with the difference that spectral dimension is replaced with temporal dimension. Localized convolution is not considered in the temporal dimension as the temporal depth is already low. The present study also focuses on the analysis of the performance of 3D-CNNs for MSTS image based crop classification. The 3D-CNN-v1 architecture and hyper-parameter settings are listed in Table 6.5.

6.5.4. RNN model for MSTS images based crop classification

A conventional RNN model is considered here for analysis. The input to this RNN model namely RNN-v1 is a 7×12 matrix where 7 indicates the temporal depth (time stamps) and 12 indicates to the spectral depth in the multisensor image data in each time stamp. Since, RNNs do not accommodate spatial contextual information, per-pixel approach is considered here. A simple LSTM unit is selected as recurrent cell. Since the temporal depth is 7, seven LSTM units constitute the R1 recurrent layer. The architecture is used in a many-to-many fashion. It means

the input to the recurrent layer is more than one dimensional (in this case, 7) and output from the recurrent layer is more than one dimensional (in this case, 7). Also, the flow of activation from each RNN cell in the recurrent layer is considered in a bidirectional manner. It means that the activation from particular recurrent cell can move to next recurrent cell (next time stamp, usual) and previous recurrent cell (previous time stamp). This helps in the crop classification and its significance is discussed in the results and discussion section later in this chapter. Further a FC layer with 20 nodes and output layer with 6 outputs is set.

The RNN-v1 architecture and hyper-parameter settings are listed in

Table 6.6. Hyperbolic tangent and softmax activation functions are used in the LSTM recurrent cells for activation and output layer respectively. A dropout strategy is considered for regularization and dropout fraction values of 0.1 and 0.1 for R1 and FC are set respectively.

Table 6.5 Architecture and hyperparameter settings of the 1D-CNN-v1, 2D-CNN-v1, and 3D-CNN-v1 models for MSTS images based crop classification.

| Model | Layer | | | |
	Input	Convolutional layer (C1)	Fully connected layer (FC)	Output layer
1D-CNN-v1	Input data size = 7×12 (time, spectral) Normalization = Batch Normalization	Filters = 32 Kernel size = 7 (time) Activation = ReLU Dropout fraction = 0.2 Pooling size = 0	Not used	Labels = 6 Activation = Softmax
3D-CNN-v1	Input data size = $7 \times 3 \times 3 \times 12$ (time, spatial, spatial, spectral) Normalization = Batch Normalization	Filters = 32 Kernel size = $7 \times 3 \times 3$ (time, spatial, spatial) Activation = ReLU Dropout fraction = 0.2 Pooling size = 0		
Training parameters	Loss function = Categorical cross entropy Learning rate = 0.01 Number of epochs = 200 Optimizer = Adam [211] 10 fold cross validation strategy is employed during training			

Table 6.6 Architecture and hyperparameter settings of the RNN-v1 model for MSTS images based crop classification.

| Model | Layer | | | |
	Input	Recurrent layer (R1)	Fully connected layer (FC)	Output layer
RNN-v1	Input data size = 7×12 (time, spectral) Normalization = Batch Normalization	Filters = 10 Kernel size = 7 (time) Activation = Tanh Dropout fraction = 0.1 Pooling size = 0 Bidirectional: Yes	Nodes= 20 Activation = ReLU Dropout fraction = 0.1 Time-distributed: Yes	Labels = 6 Activation = Softmax Time-distributed: Yes

	Input-output format: Many-to-many
Training parameters	Loss function = Categorical cross entropy Learning rate = 0.01 Number of epochs = 200 Optimizer = Adam [211] 10 fold cross validation strategy is employed during training

6.5.5. CNN-RNN models for MSTS images based crop classification

In this section, CNN and RNN architectures are used synergistically to form CNN-RNN models for crop classification with MSTS images. The previous standalone CNN or RNN models are not able to completely exploit the spectral, spatial, and temporal information simultaneously. At best, the standalone 3D-CNN model i.e. 3D-CNN-v1 is able to exploit spectral-spatial information or temporal-spatial simultaneously but not all three simultaneously. Alternatively, the standalone RNN model i.e. RNN-v1 could benefit from the temporal information but is unable to extract spatial or spectral features. Therefore, a synergistic utilization of CNN and RNN models is straightforward beneficial. In the novel models developed in the next few sections, the CNNs are used to extract spectral and/or spatial features and RNNs are used to extract temporal features. The CNN architectures are used prior to RNN architectures in a time-distributed fashion to preserve the temporal information of the MSTS images.

6.5.5.1. 1D-CNN-RNN model for MSTS images based crop classification

At first, a 1D-CNN architecture is used here to extract spectral features as the temporal information is exploited using a RNN architecture. The kernel size of the filter in the convolutional layer i.e. C1 layer is equal to the spectral depth of MSTS image i.e. 12. One dimensional spectral convolution is shown in Figure 5.4(b).

Then, a conventional RNN model is considered. Since, 1D-CNNs or RNNs cannot accommodate spatial contextual information, per-pixel approach is considered here. A simple LSTM unit is selected as recurrent cell similar to the RNN-v1 model. Since the temporal depth is 7, seven LSTM units constitute the R1 recurrent layer. Here again, the RNN architecture is used in a many-to-many fashion. It means the input to the RNN layer is more than one dimensional (in this case, 7) and output from the RNN layer is more than one dimensional (in this case, 7). Also, the flow of activation from each recurrent cell in the recurrent layer is considered in a bidirectional manner. It means that the activation from particular recurrent cell can move to next recurrent cell (next time stamp, usual) and previous RNN cell (previous time stamp). This helps in the crop classification and the significance is discussed in the results and

discussion section later in this chapter. Further a FC layer with 20 nodes and output layer with 6 outputs is set.

The consolidated CNN-RNN model is termed as 1D-CRNN-v1 hereafter in this chapter. The architecture and hyper-parameter settings are listed in Table 6.7. ReLU, hyperbolic tangent (Tanh) and softmax activation functions are used in the convolutional & FC layers, LSTM recurrent cells, and output layer respectively. A dropout strategy is considered for regularization and dropout fraction values of 0.2, 0.1 and 0.1 for C1, R1 and FC are set respectively.

Table 6.7 Architecture and hyperparameter settings of CNN-RNN based models for MSTS images based crop classification.

Model	Layer				
	Input	Convolutional layer (C1)	Recurrent layer (R1)	Fully connected layer (FC)	Output layer
1D-CRNN-v1	Input data size = 7×12 (time, spectral) Normalization: Batch Normalization	Filters = 32 Kernel size = 12 (spectral) Activation = ReLU Dropout fraction = 0.2 Pooling size = 0 Time-distributed: Yes	Filters = 10 Kernel size = 7 (time) Activation = Tanh Dropout fraction = 0.2 Pooling size = 0 Bidirectional: Yes Input-output format: Many-to-many	Nodes= 20 Activation = ReLU Dropout fraction = 0.1 Time-distributed: Yes	Labels = 6 Activation = Softmax Time-distributed: Yes
2D-CRNN-v1	Input data size = $7 \times 3 \times 3 \times 12$ (time, spatial, spatial, spectral) Normalization: Batch Normalization	Filters = 32 Kernel size = 3×3 (spatial) Activation = ReLU Dropout fraction = 0.2 Pooling size = 0 Time-distributed: Yes	Filters = 10 Kernel size = 7 (time) Activation = Tanh Dropout fraction = 0.2 Pooling size = 0 Bidirectional: Yes Input-output format: Many-to-many	Nodes= 20 Activation = ReLU Dropout fraction = 0.1 Time-distributed: Yes	Labels = 6 Activation = Softmax Time-distributed: Ye
3D-CRNN-v1	Input data size = $7 \times 3 \times 3 \times 12$ (time, spatial, spatial, spectral) Normalization: Batch Normalization	Filters = 32 Kernel size = $3 \times 3 \times 12$ (spatial, spatial, spectral) Activation = ReLU Dropout fraction = 0.2 Pooling size = 0 Time-distributed: Yes	Filters = 10 Kernel size = 7 (time) Activation = Tanh Dropout fraction = 0.2 Pooling size = 0 Bidirectional: Yes Input-output format: Many-to-many	Nodes= 20 Activation = ReLU Dropout fraction = 0.1 Time-distributed: Yes	Labels = 6 Activation = Softmax Time-distributed: Yes
Training parameters	Loss function = Categorical cross entropy Learning rate = 0.01 Number of epochs = 200 Optimizer = Adam [211] 10 fold cross validation strategy is employed during training				

6.5.5.2. 2D-CNN-RNN model for MSTS images based crop classification

The 2D-CNN architecture is used here to extract spatial features as the temporal features are capture by the RNN architecture. Since, 2D-CNNs can accommodate spatial contextual information, patch based approach is considered here. The kernel size of the filter in the convolutional layer i.e. C1 layer is a 3×3 window. The 2D-CNN architecture is employed in a time-distributed fashion in order to preserve the temporal information. Then, a conventional RNN model is considered which is identical to the RNN model used in 1D-CRNN-v1. This consolidate CNN-RNN model is termed as 2D-CRNN-v1 hereafter in this chapter. The architecture and hyper-parameter settings are listed in Table 6.7.

6.5.5.3. 3D-CNN-RNN model for MSTS images based crop classification

The 3D-CNN architecture is used here to extract spectral-spatial features as the temporal features are capture by the RNN architecture. Since, 3D-CNNs can also accommodate spatial contextual information, patch based approach is considered here. The kernel size of the filter in the convolutional layer i.e. C1 layer is a $3 \times 3 \times 12$ window. The 3D-CNN architecture is employed in a time-distributed fashion in order to preserve the temporal information.

After the convolutional layer, a conventional RNN model is considered similar to the RNN models used in 1D-CRNN-v1 and 2D-CRNN-v1. This consolidate CNN-RNN model is termed as 3D-CRNN-v1 hereafter in this chapter. The architecture and hyper-parameter settings are listed in Table 6.7.

6.6. The Concept of Permuted Local Spectral Convolutions and its Synergistic Use with RNNs for MSTS images based Crop Classification

The concept of permuted local spectral convolution is introduced and discussed in chapter 5 (refer section 5.5). Chapter 5 also establishes the significance and advantage of permuted local spectral convolutions over the conventional spectral convolutions or local spectral convolutions. In this chapter, the concept is analysed in presence of two rather one complementary information i.e. spatial and temporal. Usually, the DNN models developed for high dimensional time series image data, focus on temporal and/or spatial information more than the spectral information present. However, in context to satellite based earth observation images, the spectral information is as important as the spatial and temporal or even more. Therefore, the idea of utilizing the spectral information more efficiently along with the spatial and temporal information motivates to development of novel DNN models.

144

Therefore, in this chapter, the study and analysis of the performance of synergistic use of CNNs with permuted local spectral convolution and RNNs for MSTS based crop classification is carried out. It is postulated here that the RNNs might take advantage of permuted local spectral convolutions in MSTS image applications. The more generalized features obtained using the permuted local spectral convolutions are more abstract in nature and the CNN-RNN models built on top of this concept might improve the performance of these models.

6.7. Sub Task 2: Novel CNN and RNN Models with Permuted Local Spectral Convolutions for MSTS Image based Crop Classification

In this section, various novel CNN and RNN models are developed for MSTS image based crop classification which use the concept explained in the previous section. This section provides information about the architecture, and training parameters of the novel models developed.

6.7.1. 1D-CNN model with permuted local spectral convolution for MSTS Images based crop classification

In this 1D-CNN model namely Perm-1D-CNN-v1, first, a permutation layer which generates 10 permuted versions of spectral bands as explained in chapter 5 (refer section 5.5). Then, the 10 permuted versions are stacked together to form an ensembled spectral band stack. This stack of bands (120, 10 permutations with 12 bands in each permutation set) is used as input in the further models.

After the ensembled input is generated, one dimensional convolution layers are used. Since, the temporal depth is unaltered, one dimensional convolution in the temporal dimension will yield similar results as 1D-CNN-v1, it is insignificant to replicate the model. However, the spectral depth has now increased from 12 to 120, local spectral convolutions should yield more and may be different spectral features. Therefore, a series of spectral convolutions are performed over the spectral dimension with the help of one dimensional convolutional filters.

Three convolutional layers in time-distributed fashion are employed. The kernel size of the filters in the three convolutional layers i.e. C1, C2, and C3 layers are set to 7, 21, and 63 respectively. The kernel sizes are set based on experiments. Then, a FC layer with 50 nodes is used. Finally, an output layer with 6 classes is used. The FC and the output layer are also employed in time-distributed fashion.

The Perm-1D-CNN-v1 architecture and hyper-parameter settings are listed in Table 6.8. Rectified linear unit (ReLU) and softmax activation functions are used in the convolutional & FC layers and output layers respectively. A dropout strategy is considered for regularization and dropout fraction values of 0.2 for C1, C2, and C3 are set. A dropout fraction value of 0.1 is set for the FC layer. The loss function used is the categorical cross entropy function. It optimizes the classification accuracy based on probabilities of each class during classification. The learning rate and the number of epochs are set to 0.01 and 200 respectively.

6.7.2. 2D-CNN model with permuted input for MSTS images based crop classification

Since the input MSTS image data has 1 spectral, 2 spatial and 1 temporal dimension, it is not possible for 2D-CNNs to process this 4-dimensional data simultaneously and directly. Therefore, the permuted local spectral convolution version of the 2D-CNN model is not explored further.

6.7.3. 3D-CNN model with permuted local spectral convolution for MSTS images based crop classification

The 3D-CNN model proposed here namely Perm-3D-CNN-v1 operates window based convolution along the spectral and spatial dimensions in a time distributed fashion.

The first layer is a permutation layer similar to the Perm-1D-CNN-v1 with the difference in the spatial dimension. The Perm-3D-CNN-v1 uses a 3×3 window in the spatial dimensions input in contrast to Perm-1D-CNN-v1 which uses per-pixel approach. After the ensembled input is generated, a series of three dimensional convolutions are employed in a time distributed fashion. The three dimensional convolutions operate over the spectral and spatial dimensions.

Then, three convolutional layers are employed. The kernel size of the filters in the three convolutional layers i.e. C1, C2, and C3 layers are set to $3\times3\times7$, $1\times1\times21$, and $1\times1\times63$ respectively where 3×3 represents the spatial extents and 7, 21, or 63 represents the spectral extents of the convolutional filters. The kernel sizes are set based on experiments. Then, a FC layer with 50 nodes is used. Finally, an output layer with 6 classes is used. The FC and the output layer are also employed in time-distributed fashion. The Perm-3D-CNN-v1 architecture and hyper-parameter settings are listed in Table 6.8.

Table 6.8 Architecture and hyperparameter settings of CNN models with permuted spectral convolution for MSTS images based crop classification.

Model	Layer				
	Input	Permutation layer	Convolutional layer (C1, C2, C3)	Fully connected layer (FC)	Output layer
Perm-1D-CNN-v1	Input data size = 7×12 (time, spectral) Normalization = Batch Normalization	Permutated sets = 10 Kernel size = 12 (spectral) Activation = Linear Time-distributed: Yes	Filters = 20, 40, 60 Kernel size = 7, 21, 63 (spectral) Activation = ReLU Dropout fraction = 0.2 Time-distributed: Yes	Nodes= 50 Activation = ReLU Dropout fraction = 0.1 Time-distributed: Yes	Labels = 6 Activation = Softmax
Perm-3D-CNN-v1	Input data size = $7\times3\times3\times12$ (time, spatial, spatial, spectral) Normalization = Batch Normalization	Permutated sets = 10 Kernel size = 12 (spectral) Activation = Linear Time-distributed: Yes	Filters = 20, 30, 40 Kernel size = $3\times3\times7$, $1\times1\times21$, $1\times1\times63$ (spatial, spatial, spectral) Activation = ReLU Dropout fraction = 0.2 Time-distributed: Yes		
Training parameters	Loss function = Categorical cross entropy Learning rate = 0.01 Number of epochs = 200 Optimizer = Adam 10-fold cross validation strategy is employed during training				

6.7.4. RNN model with permuted input for MSTS images based crop classification

Permuted RNN model is similar to the conventional RNN model (RNN-v1) discussed in section 6.5.4 apart from the addition of a permutation layer. The permutation layer prepares an ensemble of permuted spectral bands similar to as in Perm-1D-CNN-v1 or Perm-3D-CNN-v1., Therefore, the input to this RNN model namely Perm-RNN-v1 is a 7×120 matrix where 7 indicates the length of time dimension (time stamps) and 120 indicates to the permuted and ensembled spectral depth in data in each time stamp. Since, RNNs do not accommodate spatial contextual information, per-pixel approach is considered here. A simple LSTM unit is selected as recurrent cell. Similar to RNN-v1, seven LSTM units constitute the R1 recurrent layer and the architecture is used in a many-to-many fashion. It means the input to the recurrent layer is more than one dimensional (in this case, 7) and output from the recurrent layer is more than one dimensional (in this case, 7). Also, the flow of activation from each recurrent cell in the recurrent layer is considered in a bidirectional manner. It means that the activation from particular recurrent cell can move to next recurrent cell (next time stamp, usual) and previous recurrent cell (previous time stamp). This helps in the crop classification and the significance is discussed in the results and discussion section later in this chapter. Further a FC layer with 20 nodes and an output layer with 6 outputs is set. The Perm-RNN-v1 architecture are listed in Table 6.9. The hyper-parameter settings are similar to RNN-v1 are also listed in Table 6.9.

Table 6.9 Architecture and hyperparameter settings of RNN model with permuted input for MSTS images based crop classification.

| Model | Layer | | | | |
	Input	Permutation layer	Recurrent layer (R1)	Fully connected layer (FC)	Output layer
Perm-RNN-v1	Input data size = 7×12 (time, spectral) Normalization = Batch Normalization	Permutated sets = 10 Kernel size = 12 (spectral) Activation = Linear Time-distributed: Yes	Filters = 10 Kernel size = $_7$ (time) Activation = Tanh Dropout fraction = 0.1 Pooling size = 0 Bidirectional: Yes Input-output format: Many-to-many	Nodes= 20 Activation = ReLU Dropout fraction = 0.1 Time-distributed: Yes	Labels = 6 Activation = Softmax Time-distributed: Yes
Training parameters	Loss function = Categorical cross entropy Learning rate = 0.01 Number of epochs = 200 Optimizer = Adam [211] 10 fold cross validation strategy is employed during training				

6.7.5. CNN-RNN models with permuted local spectral convolution for MSTS images based crop classification

In this section, CNN and RNN architectures are used synergistically to form a novel CNN-RNN model for crop classification with MSTS satellite images. These CNN-RNN models are similar to CNN-RNN models discussed in section 6.5.5 with the difference of addition of a permuted layer. The ensemble of permuted sets of spectral bands help emphasize the spectral characteristic during MSTS image based classification as discussed in section 6.6.

6.7.5.1. 1D-CNN-RNN model with permuted local spectral convolution for MSTS images based crop classification

In this 1D-CNN-RNN model namely Perm-1D-CRNN-v1, first, a permutation layer which generates 10 permuted versions of spectral bands. Then, the 10 permuted versions are stacked together to form an ensembled spectral band stack. This stack of bands (120, 10 permutations with 12 bands in each permutation set) is used as input.

After the ensembled input is generated, one dimensional convolution layers are used. Since, the temporal depth is unaltered, one dimensional convolution in the temporal dimension will yield similar results as 1D-CRNN-v1, it is insignificant to replicate the model. However, the spectral depth has now increased from 12 to 120, local spectral convolutions should yield more and may be different spectral features. Therefore, a series of spectral convolutions are performed over the spectral dimension with the help of one dimensional convolutional filters.

Three convolutional layers in time-distributed fashion are employed. The kernel size of the filters in the three convolutional layers i.e. C1, C2, and C3 layers are set to 7, 21, and 63 respectively. The kernel sizes are set based on experiments.

Then, a conventional RNN model is considered. Since, 1D-CNNs or RNNs cannot accommodate spatial contextual information, per-pixel approach is considered here. A simple LSTM unit is selected as recurrent cell. Since the temporal depth is 7, seven LSTM units constitute the R1 recurrent layer. Here again, the RNN architecture is used in a many-to-many fashion. Also, the flow of activation from each recurrent cell in the recurrent layer is considered in a bidirectional manner. Then, a FC layer with 20 nodes is used. Finally, an output layer with 6 classes is used. The FC and the output layer are also employed in time-distributed fashion.

The architecture and hyper-parameter settings are listed in Table 6.10. ReLU, hyperbolic tangent (Tanh) and softmax activation functions are used in the convolutional & FC, LSTM recurrent cells, and output layer respectively. A dropout strategy is considered for regularization and dropout fraction values of 0.2, 0.1 and 0.1 for (C1, C2, C3), R1 and FC are set respectively.

The loss function used is the categorical cross entropy function. It optimizes the classification accuracy based on probabilities of each class during classification. The learning rate and the number of epochs are set to 0.01 and 200 respectively.

6.7.5.2. 2D-CNN-RNN model with permuted input for MSTS images based crop classification

In this 2D-CNN-RNN model namely Perm-2D-CRNN-v1, first, a permutation layer which generates 10 permuted versions of spectral bands. Then, the 10 permuted versions are stacked together to form an ensembled spectral band stack. This stack of bands (120, 10 permutations with 12 bands in each permutation set) is used as input.

After the ensembled input is generated, a two dimensional convolution layer is used. Since, two dimensional convolutions can accommodate spatial contextual information, patch based convolution approach is considered here. The kernel size of the filter in the convolutional layer i.e. C1 layer is a 3×3 window. The convolutional layer is employed in time-distributed fashion.

Then, a conventional RNN model is considered identical to the RNN model used in Perm-1D-CRNN-v1. The architecture of the Perm-2D-CRNN-v1 is listed in Table 6.10. The hyper-parameter settings are similar to the Perm-1D-CRNN-v1 model and are listed in Table 6.10.

6.7.5.3. 3D-CNN-RNN model with permuted local spectral convolution for MSTS images based crop classification

In this 3D-CNN-RNN model namely Perm-3D-CRNN-v1, first, a permutation layer which generates 10 permuted versions of spectral bands. Then, the 10 permuted versions are stacked together to form an ensembled spectral band stack. This stack of bands (120, 10 permutations with 12 bands in each permutation set) is used as input.

After the ensembled input is generated, a series of three dimensional convolution layers are used. The three dimensional convolution layers are used here to extract spectral-spatial features as the temporal features are capture by the RNN architecture. Since, three dimensional convolutions can also accommodate spatial contextual information, patch based convolution approach is considered here. The kernel size of the filters in the three convolutional layers i.e. C1, C2, and C3 layers are set to $3\times3\times7$, $1\times1\times21$, and $1\times1\times63$ respectively. The kernel sizes are set based on experiments.

Then, a conventional RNN model is considered identical to the RNN model used in Perm-1D-CRNN-v1 and Perm-2D-CRNN-v1. The architecture of the Perm-3D-CRNN-v1 is listed in Table 6.10. The hyper-parameter settings are similar to the Perm-3D-CRNN-v1 model and are listed in Table 6.10.

Table 6.10 Architecture and hyperparameter settings of CNN-RNN with permuted spectral convolution based models for MSTS images based crop classification.

Model	Input	Permutation layer	Convolutional layer 1 (C1, C2, C3)	Recurrent layer (R1)	Fully connected layer (FC)	Output layer
			Layer			
Perm-1D-CRNN-v1	Input data size = 7×12 (time, spectral) Normalization: Batch Normalization	Permutated sets = 10 Kernel size = 12 (spectral) Activation = Linear Time-distributed: Yes	Filters = 20, 40, 60 Kernel size = 7, 21, 63 (spectral) Activation = ReLU Dropout fraction = 0.2 Time-distributed: Yes	Filters = 10 Kernel size = $_7$ (time) Activation = Tanh Dropout fraction = 0.2 Pooling size = 0 Bidirectional: Yes Input-output format: Many-to-many	Nodes= 20 Activation = ReLU Dropout fraction = 0.1 Time-distributed: Yes	Labels = 6 Activation = Softmax Time-distributed: Yes
Perm-2D-CRNN-v1	Input data size = $7\times3\times3\times12$ (time, spatial, spatial, spectral) Normalization: Batch Normalization	Permutated sets = 10 Kernel size = 12 (spectral) Activation = Linear Time-distributed: Yes	Filters = 32 Kernel size = 3×3 (spatial, spatial) Activation = ReLU Dropout fraction = 0.2 Time-distributed: Yes	Filters = 10 Kernel size = $_7$ (time) Activation = Tanh Dropout fraction = 0.2 Pooling size = 0 Bidirectional: Yes Input-output format: Many-to-many	Nodes= 20 Activation = ReLU Dropout fraction = 0.1 Time-distributed: Yes	Labels = 6 Activation = Softmax Time-distributed: Ye
Perm-3D-CRNN-v1	Input data size = $7\times3\times3\times12$ (time, spatial, spatial, spectral) Normalization: Batch Normalization	Permutated sets = 10 Kernel size = 12 (spectral) Activation = Linear Time-distributed: Yes	Filters = 20, 30, 40 Kernel size = $3\times3\times7$, $1\times1\times21$, $1\times1\times63$ (spatial, spatial, spectral) Activation = ReLU Dropout fraction = 0.2 Time-distributed: Yes	Filters = 10 Kernel size = $_7$ (time) Activation = Tanh Dropout fraction = 0.2 Pooling size = 0 Bidirectional: Yes Input-output format: Many-to-many	Nodes= 20 Activation = ReLU Dropout fraction = 0.1 Time-distributed: Yes	Labels = 6 Activation = Softmax Time-distributed: Yes
Training parameters	Loss function = Categorical cross entropy Learning rate = 0.01 Number of epochs = 200 Optimizer = Adam [211] 10 fold cross validation strategy is employed during training					

6.8. Results and Discussion

A. *Quantitative analysis*

The results from the models presented in section 6.5, and 6.7 are discussed in this section. Both, qualitative and quantitative assessment are carried out. Overall, results from 12 different models are compared and discussed here. These are 1D-CNN-v1, 3D-CNN-v1, RNN-v1, 1D-CRNN-v1, 2D-CRNN-v1, 3D-CRNN-v1, Perm-1D-CNN-v1, Perm-3D-CNN-v1, Perm-RNN-v1, Perm-1D-CRNN-v1, Perm-2D-CRNN-v1, and Perm-3D-CRNN-v1. Figure 6.10 show classification results obtained using these 12 models. Also, Table 6.11 shows quantitative assessment of the classification performance of these 12 models. It is evident from Table 6.11 that Perm-1D-CNN-v1 shows the best classification accuracy of 99.70%. However, most of the other models also show classification accuracies of the same order i.e. above 90%. Therefore, it is difficult to recognize the best model with only quantitative assessment. A detailed qualitative assessment is carried out with the help of ground truth information.

Table 6.11 Summary of the classification performance of the compared classifiers.

Method	Overall accuracy (%)
1D-CNN-v1	95.66
3D-CNN-v1	97.45
RNN-v1	87.29
1D-CRNN-v1	89.70
2D-CRNN-v1	94.20
3D-CRNN-v1	96.00
Perm-1D-CNN-v1	99.70
Perm-3D-CNN-v1	99.15
Perm-RNN-v1	94.73
Perm-1D-CRNN-v1	96.85
Perm-2D-CRNN-v1	95.98
Perm-3D-CRNN-v1	96.00
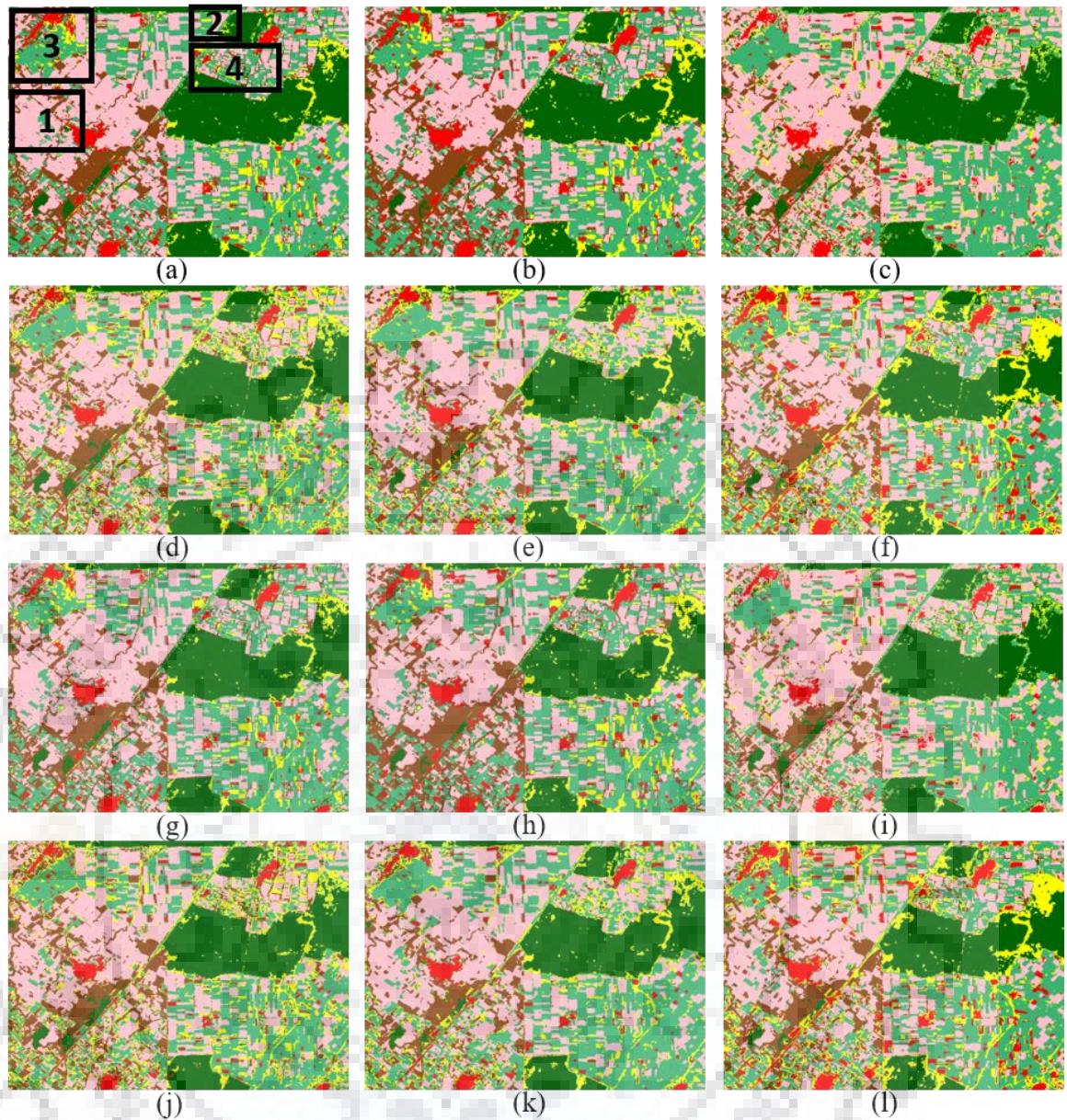
Colour scheme: Wheat | Built-up | Sugarcane | Grassland | Forest | Popular

Figure 6.10 Classified image; (a) 1D-CNN-v1. Markers; 1: GTI-1, 2: GTI-2, 3: GTI-3, and 4: GTI-4. (b) 3D-CNN-v1, (c) RNN-v1, (d) 1D-CRNN-v1, (e) 2D-CRNN-v1, (f) 3D-CRNN-v1, (g) Perm-1D-CNN-v1, (h) Perm-3D-CNN-v1, (i) Perm-RNN-v1, (j) Perm-1D-CRNN-v1, (k) Perm-2D-CRNN-v1, and (l) Perm-3D-CRNN-v1.

B. *Qualitative assessment*

A set of images collected during the crop growing season are used as ground truth to visually evaluate the performance of these models. The visual interpretation is an evaluation of the generalization performance of the models. Few of these ground truth images are discussed here. The subset areas selected for visual inspection are marked in Figure 6.6, Figure 6.7, and Figure 6.10(a). These subset areas are termed as ground truth images (GTI). Figure 6.11 shows a ground truth image (GTI-1) which is patch from the original study area (marked as 1 in Figure 6.6). The GTI-1 has two components; the first component is a GTI from November, 2017 and

the second component is GTI from February, 2018. It can be observed from GTI of November, 2017 that no major crop is growing in the area and the green patches and tree lines are popular plantations (highlighted in yellow box). Whereas, in the GTI of February, 2018, wheat crop is grown in the area (sample highlighted in green box). Visual interpretation of the corresponding patches from the classified images of the compared models indicate that all the models have correctly classified the crop as wheat. However, it is critical to observe that though Perm-1D-CNN-v1 or Perm-3D-CNN-v1 does not perform temporal convolution, they able to correctly classify the image patch. This indicates to the ability of CNNs to focus on any particular time stamp based on training labels and provide classification. Even though, Perm-1D-CNN-v1 show highest classification accuracy, it inability to not consider temporal pattern makes it less generalized model in MSTS image based crop classification.

Figure 6.12 shows a ground truth image (GTI-2) which is patch 2 from the original study area. The GTI-2 also has two components; the first component is a GTI from November 2017 and the second component is GTI from February, 2018. It can be observed from GTI of November, 2017 that major crop growing in the area is sugarcane (highlighted in yellow box) and some fallow lands are present. Whereas, in the GTI of February, 2018 that wheat crop is grown in these fallow lands (highlighted in green box) and most of the sugarcane is harvested which in turn leaves new fallow lands. Visual interpretation of the corresponding patches from the classified images of the compared models reveal peculiar information.

Figure 6.11 Visual interpretation and evaluation of classification performance of the compared classifiers for GTI-1.

Models with only convolutional layers and no recurrent layers such as 1D-CNN-v1 or 3D-CNN-v1 focus on a particular time stamp and does not exploit the temporal pattern present in the MSTS data. For example, in Figure 6.12, 3D-CNN-v1 (in both, standard and permuted input case) classifies the temporary fallow lands (unused land in between two crop cycles) as grasslands (shown in yellow). On the contrary, models with recurrent layers such as RNN-v1 or 3D-CRNN-v1 are able to ignore the temporary fallow lands and classified the areas as either sugarcane or wheat. It is important to emphasize here that even though the Perm-1D-CNN-v1 provides highest classification accuracy, it is unable to capture the temporal pattern in the MSTS data. Therefore, it is obvious to infer those standalone convolutional models are overfitting.

It can also be observed from Figure 6.12 that all the standard input based models have misclassified few sugarcane fields (highlighted in red box) as built-up due to dominance of HH backscatter response in the spectral dimension. Whereas, the Perm-RNN-v1 and Perm-3D-CRNN-v1 models have correctly classified the same sugarcane fields. This indicates that models such as Perm-RNN-v1 or Perm-3D-CRNN-v1 are able to capture the spectral attributes more efficiently because of the training of these models with ensemble of permuted spectral bands (averaging out the dominance of any particular spectral band). This novel approach of permuted spectral stacking improved the generalization ability of these models and complemented the temporal attribute of the crops.
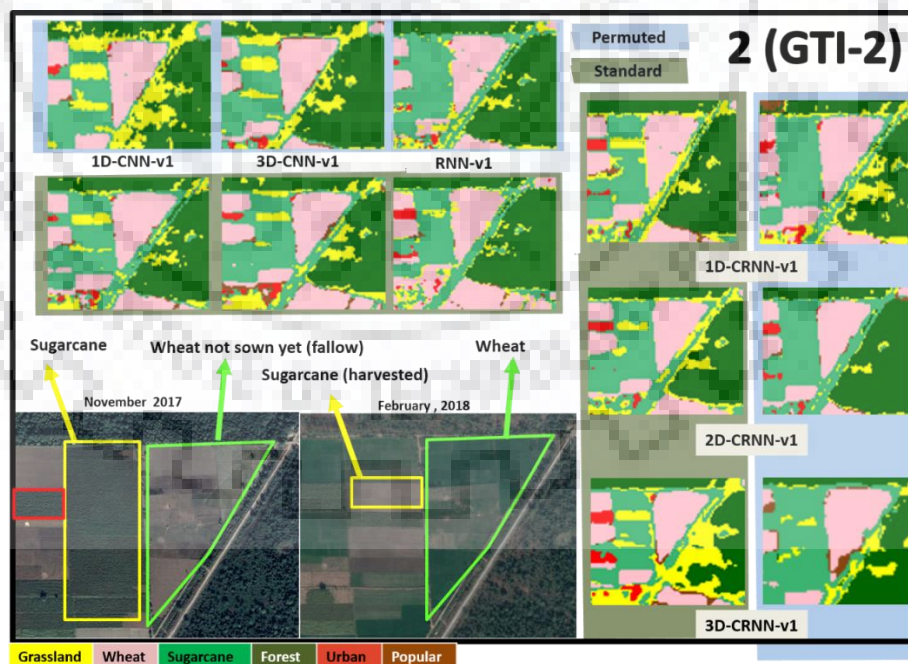
Figure 6.12 Visual interpretation and evaluation of classification performance of the compared classifiers for GTI-2.

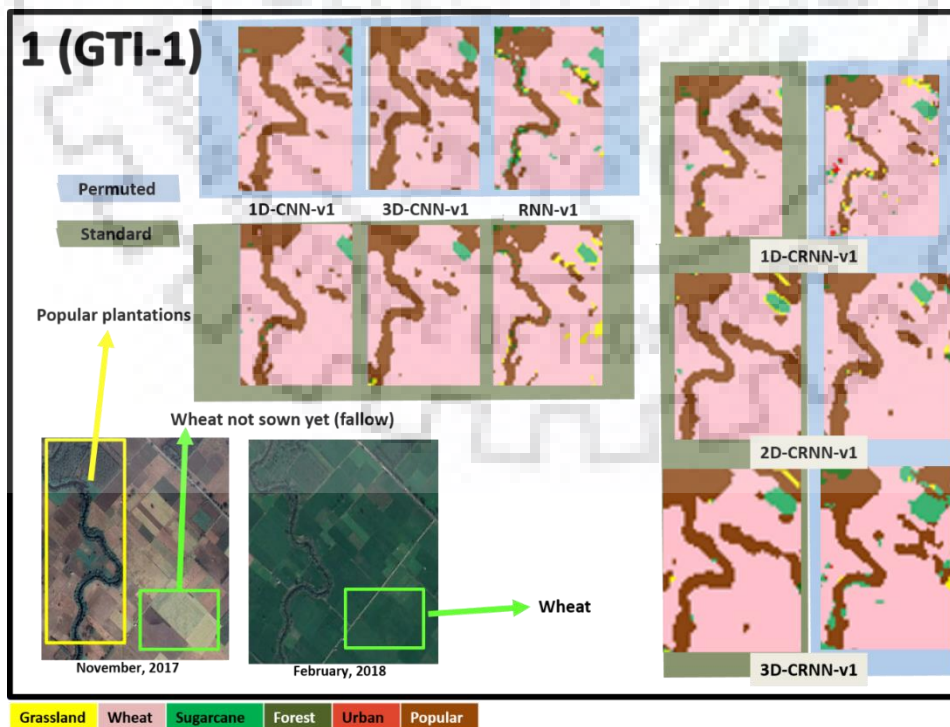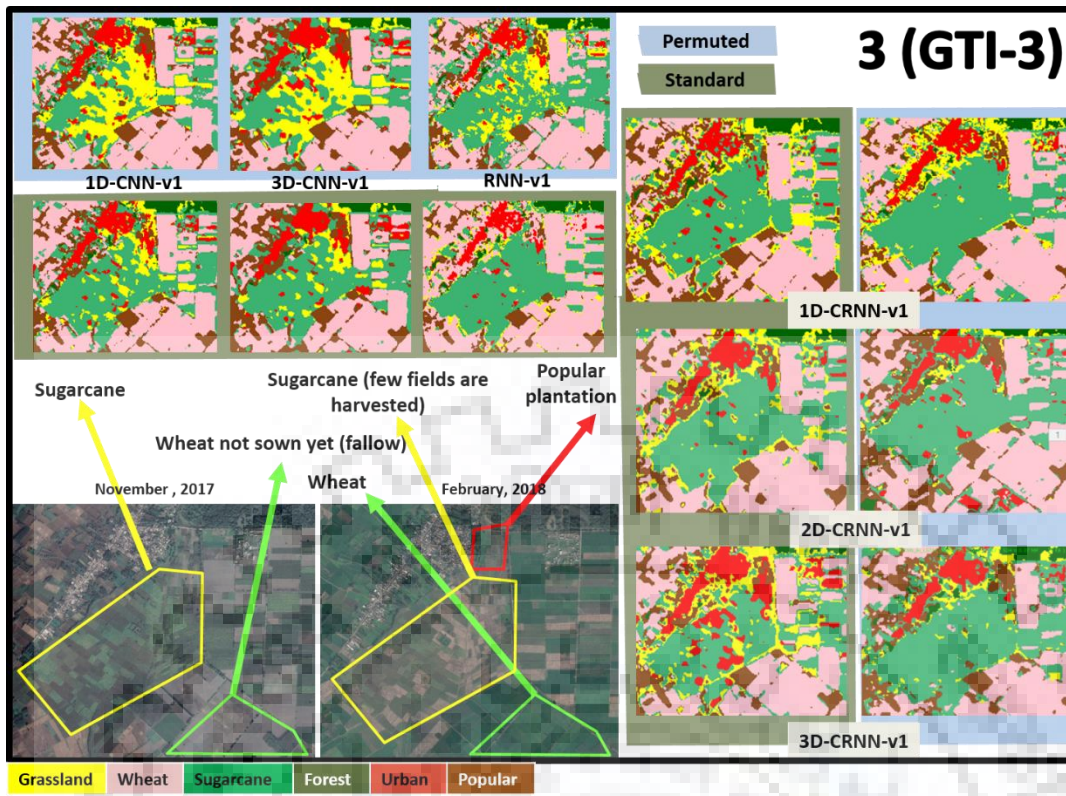Figure 6.13 Visual interpretation and evaluation of classification performance of the compared classifiers for GTI-3.
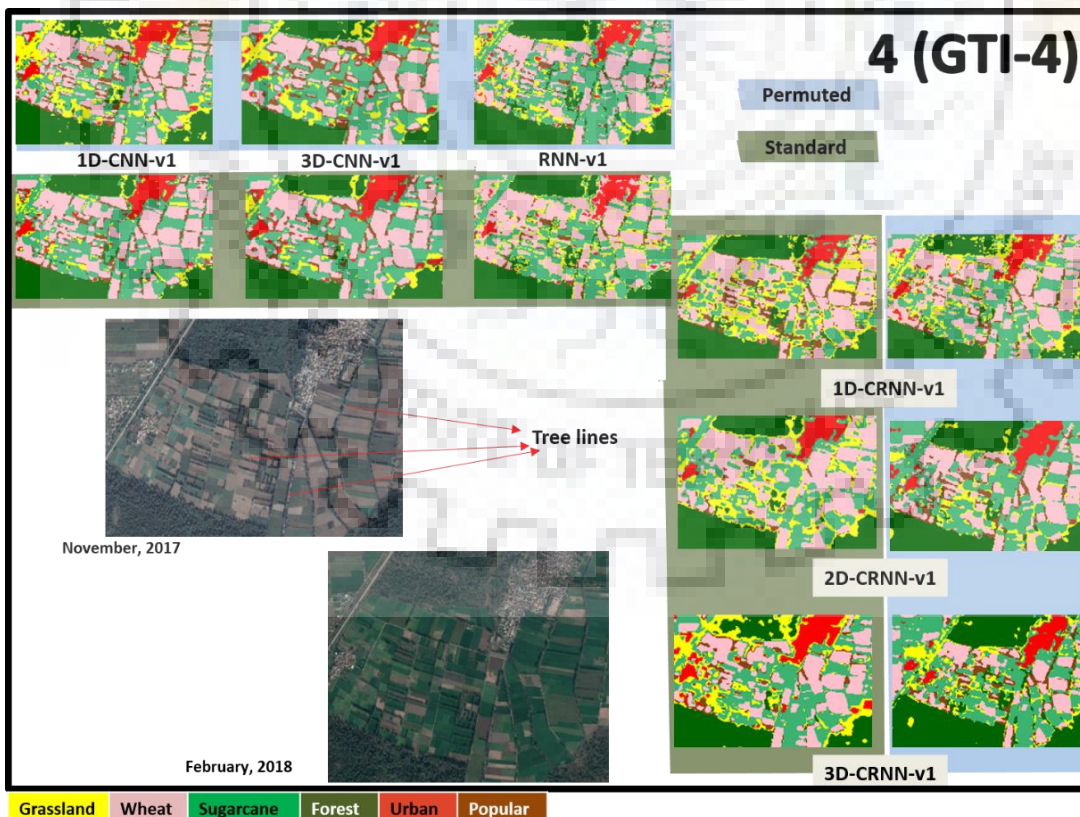
Figure 6.14 Visual interpretation and evaluation of classification performance of the compared classifiers for GTI-4.

Similarly, in Figure 6.13, the cropland highlighted in yellow box in the ground truth image (GTI-3) is a sugarcane field and is classified as sugarcane by the Perm-3D-CRNN-v1. In contrast, the Perm-1D-CNN-v1 model focused more the spectral information due to the permuted spectral band stacking than on the temporal information and hence, classified the sugarcane field on the basis of a particular time stamp and not on the basis of the entire temporal profile. This analysis also indicates to the overfitting nature of convolutional only models in case of MSTS images based crop classification. The efficient exploitation of the spectral, spatial, and temporal information present in the MSTS images by the Perm-3D-CRNN-v1 is also evidenced from the correct classification of popular plantations in the subset image (highlighted in red box). On the other hand, Perm-1D-CNN-v1misclassifies it as built-up.

The GTI in Figure 6.14 (GTI-4) shows crop fields with lots of tree lines in between (highlighted in red lines) and forests. Once again, an efficient exploitation of the spectral, spatial, and temporal information in the MSTS images is done by the Perm-3D-CRNN-v1 model as it retains the spectral and structural identities of tree line present in between crop field while suppressing the temporary fallow lands spectral attributes at the same time.

In summary, the visual interpretations highlight the overfitting nature of the otherwise quantitatively the best classification model i.e. Perm-1D-CNN-v1. Also, these interpretations highlight the good generalization ability of the proposed Perm-3D-CRNN-v1 model which is an important attribute of a classifier. It is important here to mention that the overall classification accuracy of the Perm-3D-CRNN-v1 model (96.00%) is only 3.7% less than the Perm-1D-CNN-v1 (99.70%). Hence, it is not obnoxious to assume that the Perm-3D-CRNN-v1 model is the best classifier among all the developed and compared classifiers.

6.9. Conclusion

This chapter presents a comprehensive study and analysis of the performance of conventional CNN and RNN models in MSTS images based crop classification. The analysis indicated to satisfactory performance of conventional CNN and RNN models in MSTS based crop classification. Also, preferences given to temporal and spatial characteristics of MSTS images by the conventional DNN models is observed. This led to underutilization of spectral characteristics of MSTS images and therefore resulted in less generalized crop classification models. To address this issue, novel CNN and RNN models are proposed and developed on the concept of local spectral convolutions on ensemble stack of permuted spectral bands of MSTS images. The local spectral convolutions in an ensembled spectral dimension improves the generalization ability of DNN models and also complements the spatial and temporal features. Overall, 12 CNN and RNN models are developed during the study namely 1D-CNN-v1, 3D-

CNN-v1, RNN-v1, 1D-CRNN-v1, 2D-CRNN-v1, 3D-CRNN-v1, Perm-1D-CNN-v1, Perm-3D-CNN-v1, Perm-RNN-v1, Perm-1D-CRNN-v1, Perm-2D-CRNN-v1, and Perm-3D-CRNN-v1.

These models are trained on MSTS satellite images of areas having crop fields in challenging landscape. Wheat and sugarcane are the major crops practices in the area considered here for study. The area, apart from croplands, also contain, forests, built-up, grasslands, and deciduous plantations. The study area is a good representation of real world crop classification scenarios particularly in India. The performance of these models in this study area presents be good indicator of their generalization ability.

Quantitative analysis of the classification results from these models indicate to superior performance of the Perm-1D-CNN-v1 model with classification accuracy of 99.7%. However, the Perm-3D-CRNN-v1, which provides 96.00% classification accuracy, indicate to have higher generalization ability during qualitative and visual assessments. The spectral and spatial attributes of land covers (e.g. built-up or popular plantations) is well captured by the Perm-3D-CRNN-v1 whereas the Perm-1D-CNN-v1 appears to be overfitting. Therefore, the Perm-3D-CRNN-v1 is the best model among the classifiers considered.

Chapter 7

CONCLUSIONS AND FUTURE SCOPE

The main objective of this thesis is the study and critical analysis of the performance of conventional CNN and RNN models and further development of the novel CNN and RNN models to exploit the spectral and/or spatial and/or temporal information present in unimodal/single senor or multimodal/multisensor satellite images. Special focus during the development of novel DNN models is given to develop models showing better generalization performance in mixed class scenarios.

Contributions of the thesis

In the attempt, four tasks are set to forward the stated objective. The tasks are based on; type(s) of satellite image(s) used, application for which the task is realized, multi-modality images involvement, and presence of time series information. The tasks are satisfactorily completed and are presented in four chapters i.e. chapter 3, 4, 5, and 6.

In *Chapter 3,* a study where an attempt is made to extract information from one of the most popular satellite image modality i.e. PolSAR images is presented. The study is composed of three sub-tasks.

Under the first sub task i.e. *sub task 1*, the investigation of polarization signatures (PSs) as potential PolSAR features for land cover (LC) classification in mixed class scenarios is done. Novel PSs based features i.e. PSCFs are computed based on the degree of similarity between observed and canonical PSs. Further, PSCFs are analysed, and are compared with decomposition based features for the purpose of LC classification. Comparisons indicate potential of PSCFs as significant features for LC classification in mixed land cover class scenarios.

Second sub task i.e. *sub task 2*, covers successful development and implementation of an optimal class boundary estimator based DT classifier which exploits the PSCFs to achieve improved LC classification. During the development of the optimal class boundary estimator based DT classifier, image and individual class-wise PSCFs statistics are utilized to formulate empirical relationship between classification performance indicator (overall accuracy) and class boundary values. Optimization of this empirical relationship with genetic algorithm resulted in optimal class boundaries which are utilized in a DT to perform LC classification. Classification results using this approach are compared with classification results from RBF-SVM, while keeping the features identical i.e. PSCFs. Classification performance of the proposed approach especially user and producer accuracies for built-up class (72.4% and 71.0% respectively) and

tall veg class (73.8%, and 89.1% respectively) indicate that the proposed approach, i.e. the optimal class boundary estimator based DT classifier, achieved better classification results in mixed LC class scenarios.

In the third sub task i.e. *sub task 3*, development of a two dimensional CNN (2D-CNN) based novel model for LC classification is presented. The 2D-CNNs can directly exploit PSs to perform LC classification, and therefore, the approach reduces the efforts to compute PSCFs. Also, since 2D-CNNs are powerful image extractors, they are able to extract the subtle contextual patterns present in PSs of different LCs. In summary, a novel PSs-CNN LC classification model is realized. Classification results from this PSs-CNN model are compared with NED and NSCM based minimum distance classifiers which are popularly used with PSs for land cover classification. Comparison indicated a superior performance of the novel PSs-CNN model for land cover classification.

Comparison between the two novel approaches indicates that the optimal class boundary estimator based DT classifier is showing an approx. 3% improved classification accuracy than PSs-CNN classifier. However, the PSs-CNN model is easier to implement, does not require an extra step (computing PSCFs), and is scalable.

Overall, the study concludes that; i) PSs have shown significant potential as PolSAR features for land cover applications, ii) segregation of built-up class and tall vegetation class in PolSAR data can be improved by using optimal class boundary estimators with decision trees for classification, and iii) 2D-CNNs have shown potential as classifiers with PSs for LC classification in terms of ease of implementation, and scalability.

Chapter 4, presents an efficient approach to map flood affected villages using PolSAR image. The approach included three stages i.e. –*first*, built-up area extraction, –*second*, village detection, and –*third*, flood affected village mapping.

In the first stage, the approach utilized PolSAR polarization signatures (PSs) and CNNs to extract built-up areas from a PolSAR image. A novel PSs-CNN classifier is developed. The PSs-CNN model provided improved built-up area extraction performance in contrast to the NED and the NSCM classifiers. In the second stage, the approach introduced and implemented a novel village detection methodology. The methodology uses a detection filter to detect whether a particular object in the built-up area image (obtained in stage 1) is a village or not. The detection filter uses *apriori* knowledge and image morphology to detect villages in the built-up area image. The detection methodology also significantly removes false alarms and misclassifications. In the third and final stage, flood affected villages are identified and non-flooded villages are masked out using another image morphology based methodology. This methodology computed overlap

ratios between objects (successfully detected as villages) and flooded area and utilized it for identification of flood affected villages.

Quantitative and qualitative assessment of the performance of the novel approach over two different study areas is carried out. Different artificial flood scenarios are created to evaluate the performance of developed approach. The developed approach is evaluated under these flood scenarios and flood affected villages are identified and mapped. Number of villages successfully detected and false alarms are computed with the help of ground truth information (actual number of villages or towns in the study regions are known). Results indicate to high detection accuracy (above 90%), low false alarms and true negatives. The approach is scalable and potentially satisfactory for real time usage.

Chapter 5 presents a critical analysis of the potential and performance of the one, two, and three dimensional CNNs, where dimensionality refers to the convolved features: spectral, spatial or both. These CNNs are applied to land cover classification with multisensor (PolSAR-MS) satellite image data. Two novel models are introduced, where local filters in either spectral domain (Perm-LS-CNN) or spatial-spectral domains (Perm-LSS-CNN) are applied after a permutation layer. The permutation layer generates an ensemble of band combinations, which are drawn upon by the CNN in order to generate powerful features for land cover classification. Classification results demonstrate the improved generalization ability of these models compared to both alternative CNN architectures and other machine learning models (SVM and random forests). Perm-LSS-CNN also fully uses both PolSAR and multispectral information, improving results compared to other popular methods for PolSAR-MS data based land cover classification.

A separability measure is used for the model significance evaluation. Separability analysis states that the Perm-LSS-CNN transforms the input data space into a feature space where separability among classes increases. This increase in separability is achieved due to the inclusion of permuted local spectral convolution along with the local spatial convolution which is not available in the 1D-CNN-v1, 2D-CNN-v1, or 3D-CNN-v1 CNN models. A fine analysis on ground truth data provides an extra qualitative assessment which complements the quantitative classification results. These demonstrate that the proposed Perm-LSS-CNN model is able to generalize much better while keeping high classification accuracies.

Chapter 6 presents a comprehensive study and analysis of the performance of conventional CNN and RNN models in MSTS images based crop classification. The analysis indicated to satisfactory performance of conventional CNN and RNN models in MSTS based crop classification. Also, preferences given to temporal and spatial characteristics of MSTS images by the conventional DNN models is observed. This led to underutilization of spectral

characteristics of MSTS images and therefore resulted in less generalized crop classification models. To address this issue, novel CNN and RNN models are proposed and developed on the concept of local spectral convolutions on ensemble stack of permuted spectral bands of MSTS images. The local spectral convolutions in an ensembled spectral dimension improves the generalization ability of DNN models and also complements the spatial and temporal features. Overall, 12 CNN and RNN models are developed during the study namely 1D-CNN-v1, 3D-CNN-v1, RNN-v1, 1D-CRNN-v1, 2D-CRNN-v1, 3D-CRNN-v1, Perm-1D-CNN-v1, Perm-3D-CNN-v1, Perm-RNN-v1, Perm-1D-CRNN-v1, Perm-2D-CRNN-v1, and Perm-3D-CRNN-v1.

These models are trained on MSTS satellite images of areas having crop fields in challenging landscape. Wheat and sugarcane are the major crops practices in the area considered here for study. The area, apart from croplands, also contain, forests, built-up, grasslands, and deciduous plantations. The study area is a good representation of real world crop classification scenarios particularly in India. The performance of these models in this study area presents be good indicator of their generalization ability.

Quantitative analysis of the classification results from these models indicate to superior performance of the Perm-1D-CNN-v1 model with classification accuracy of 99.7%. However, the Perm-3D-CRNN-v1, which provides 96.00% classification accuracy, indicate to have higher generalization ability during qualitative and visual assessments. The spectral and spatial attributes of land covers (e.g. built-up or popular plantations) is well captured by the Perm-3D-CRNN-v1 whereas the Perm-1D-CNN-v1 appears to be overfitting. Therefore, the Perm-3D-CRNN-v1 is the best model among the classifiers considered.

Future Scope

The research presented in this thesis has extensive scope and possibilities. A few major possibilities are:

- The PSs-CNN model presented in chapter 3 uses PSs in H-V basis. However, other basis such as in [149] may provide complementary information which can lead to improved land cover classification performance.
- In chapter 4, the approach which utilizes CNNs, *apriori* information, and image morphology to detect human settlements affected by flood disaster has potential to be developed into an operational system.
- In chapter 5, local spectral and local-spectral-spatial convolutions on ensembled input provides better classification results in terms of generalization. However, the ensembled input increases the size of the input which increases the processing time. This increase in the input space can be optimized via more efficient permutation strategies.
- The CNN and RNN models developed for MSTS images based crop classification developed in chapter 6 can also be optimized further in a manner similar to as for models in chapter 5.

A1. Kennaugh Matrix

Kennaugh matrix or K-matrix is the representation of backscattered power from target in terms of the receiving antenna. The kennaugh matrix represents the received power, given the receiving antenna polarization characteristics. It is real quantity unlike scattering matrix. It is similar to Muller matrix but explains for the radar backscattered case instead of optics. Kennaugh matrix is power matrix rather than amplitude and phase measure in scattering and coherency matrix respectively. Hence, all element of Kennaugh matrix are real quantities. Given

$$[S] = \begin{bmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{bmatrix} \tag{A.1}$$

where, the leading subscript represents the receive polarization state and the trailing subscript represents the transmit polarization state. In the monostatic case, K-matrix calculated from scattering parameters given in equation (A.1) as follows:

$$K_{11} = \frac{1}{2}\left[|S_{xx}|^2 + 2|S_{xy}|^2 + |S_{yy}|^2 \right] \tag{A.2}$$

$$K_{12} = K_{21} = \frac{1}{2}\left[|S_{xx}|^2 - |S_{yy}|^2 \right] \tag{A.3}$$

$$K_{13} = K_{31} = \mathrm{Re}\left\{ S_{xx}*\mathrm{conj}(S_{xy}) + S_{xy}*\mathrm{conj}(S_{yy}) \right\} \tag{A.4}$$

$$K_{14} = K_{41} = \mathrm{Im}\left\{ S_{xx}*\mathrm{conj}(S_{xy}) + S_{xy}*\mathrm{conj}(S_{yy}) \right\} \tag{A.5}$$

$$K_{22} = \frac{1}{2}\left[|S_{xx}|^2 - 2|S_{xy}|^2 + |S_{yy}|^2 \right] \tag{A.6}$$

$$K_{23} = K_{32} = \mathrm{Re}\left\{ S_{xx}*\mathrm{conj}(S_{xy}) - S_{xy}*\mathrm{conj}(S_{yy}) \right\} \tag{A.7}$$

$$K_{24} = k_{42} = \mathrm{Im}\left\{ S_{xx}*\mathrm{conj}(S_{xy}) - S_{xy}*\mathrm{conj}(S_{yy}) \right\} \tag{A.8}$$

$$K_{33} = |S_{xy}|^2 + \mathrm{Re}\{ S_{xx}*\mathrm{conj}(S_{yy}) \} \tag{A.9}$$

$$K_{34} = K_{43} = \mathrm{Im}\{ S_{xx}*\mathrm{conj}(S_{yy}) \} \tag{A.10}$$

$$K_{44} = |S_{xy}|^2 - \mathrm{Re}\{ S_{xx}*\mathrm{conj}(S_{yy}) \} \tag{A.11}$$

The K-matrix is represented as

$$K = \begin{bmatrix} K_{11} & K_{12} & K_{13} & K_{14} \\ K_{21} & K_{22} & K_{23} & K_{24} \\ K_{31} & K_{32} & K_{33} & K_{34} \\ K_{41} & K_{42} & K_{43} & K_{44} \end{bmatrix} \tag{A.12}$$

This K-matrix is used in the calculation of co- and cross-polarization signatures (refer equation (3.1)).

Journals

[1] G. S. Phartiyal, N. Bordu, D. Singh, H. Yahia, and K. Daoudi, 'Permuted Spectral and Permuted Spectral-Spatial CNN Models for PolSAR-Multispectral Data based Land Cover Classification', *Int. J. Remote Sens.*, vol. 42, no. 3, pp. 1096–1120, 2020.

[2] G. S. Phartiyal, K. Kumar, and D. Singh, 'An Improved Land Cover Classification Using Polarization Signatures for PALSAR 2 Data', *Adv. Sp. Res.*, vol. 65, no. 11, pp. 2622–2635, Jun. 2020.

[3] G. S. Phartiyal and D. Singh, 'Comparative Study on Deep Neural Network Models for Crop Classification Using Time Series PolSAR and Optical Data', *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 425, pp. 675–681, Nov. 2018.

Conferences

[1] G. S. Phartiyal, D. Singh, N. Brodu, and H. Yahia, 'Improved Utilization of PolSAR Polarization Signatures Using Convolutional-Deep Neural Nets for Land Cover Classification', in *International Geoscience and Remote Sensing Symposium*, 2019, pp. 5824–5827.

[2] G. S. Phartiyal, N. Brodu, D. Singh, and H. Yahia, 'A mixed spectral and spatial Convolutional Neural Network for Land Cover Classification using SAR and Optical data', in *EGU General Assembly*, 2018, vol. 20, p. 12647.

[3] G. S. Phartiyal, K. Kumar, D. Singh, and K. P. Singh, 'Optimal use of polarimetric signature on PALSAR-2 data for land cover classification', *Int. Geosci. Remote Sens. Symp.*, vol. 2017-July, pp. 4558–4561, 2017.

REFERENCES

[1] L. Samaniego, A. Bárdossy, and K. Schulz, 'Supervised classification of remotely sensed imagery using a modified k-NN technique', *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2112–2125, 2008.

[2] J. Li, J. M. Bioucas-Dias, and A. Plaza, 'Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning', *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, 2010.

[3] Q. Du and C. I. Chang, 'A linear constrained distance-based discriminant analysis for hyperspectral image classification', *Pattern Recognit.*, vol. 34, no. 2, pp. 361–373, 2001.

[4] O. A. B. Penatti, K. Nogueira, and J. A. Dos Santos, 'Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 44–51.

[5] L. Zhang, L. Zhang, and B. Du, 'Deep learning for Remote Sensing Data: A Technical Tutorial in the State of the Art', *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22–40, 2016.

[6] P. A. Townsend and S. J. Walsh, 'Modeling floodplain inundation using an integrated GIS with radar and optical remote sensing', *Geomorphology*, vol. 21, no. 3, pp. 295–312, 1998.

[7] R. P. Singh, D. R. Mishra, A. K. Sahoo, and S. Dey, 'Spatial and temporal variability of soil moisture over India using IRS P4 MSMR data', *Int. J. Remote Sens.*, vol. 26, no. 10, pp. 2241–2247, 2005.

[8] A. K. Maurya, D. Murugan, D. Singh, and K. P. Singh, 'A Step for Digital Agriculture by Estimating Near Real Time Soil Moisture with Scatsat-1 Data', in *International Geoscience and Remote Sensing Symposium*, 2019, pp. 5698–5701.

[9] G. Venkataraman, G. Singh, and Y. Yamaguchi, 'Fully polarimetric ALOS PALSAR data applications for snow and ice studies', *Int. Geosci. Remote Sens. Symp.*, pp. 1776–1779, 2010.

[10] G. Singh, Y. Yamaguchi, S. E. Park, W. M. Boerner, Y. Cui, and G. Venkataraman, 'Categorization of the glaciated terrain of Indian Himalaya using CP and FP mode SAR', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 3, pp. 872–880, 2014.

[11] R. M. Gairola, B. S. Gohil, and P. C. Pandey, 'Retrieval of precipitable water from bhaskara-II microwave measurements and its comparison with NOAA-7 and radiosonde data', *Remote Sens. Environ.*, vol. 18, no. 2, pp. 125–135, 1985.

[12] S. Sil, A. Chakraborty, S. K. Basu, and P. C. Pandey, 'Response of OceanSat II

scatterometer winds in the Bay of Bengal circulation', *Int. J. Remote Sens.*, vol. 35, no. 14, pp. 5315–5327, 2014.

[13] R. Avtar, H. Sawada, W. Takeuchi, and G. Singh, 'Characterization of forests and deforestation in Cambodia using ALOS/PALSAR observation', *Geocarto Int.*, vol. 27, no. 2, pp. 119–137, 2012.

[14] P. C. Pandey, V. K. Dadhwal, B. Sahai, and P. P. Kale, 'An optimal estimation technique for increasing the accuracy of crop forecasts by combining remotely sensed and conventional forecast', *Int. J. Remote Sens.*, vol. 13, no. 14, pp. 2735–2741, 1992.

[15] Y. Kim and J. J. van Zyl, 'A Time-Series Approach to Estimate Soil Moisture Using Polarimetric Radar Data', *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2519–2527, 2009.

[16] O. Altan, T. J. Tanzi, M. Chandra, F. Sunar, and L. Halounová, 'Spatially supported disaster management: Introduction to the special issue "gI for disaster management', *ISPRS Int. J. Geo-Information*, vol. 9, no. 5, pp. 10–11, 2020.

[17] B. S. Gohil, A. K. Mathur, and P. C. Pandey, 'An algorithm for sea surface temperature estimation from ERS-1 ATSR using moisture dependent coefficients: A simulation study', *Int. J. Remote Sens.*, vol. 15, no. 5, pp. 1161–1167, 1994.

[18] S. Basu, R. M. Gairola, C. M. Kishtawal, and P. C. Pandey, 'Empirical orthogonal function analysis of humidity profiles over the Indian Ocean and an assessment of their retrivability using satellite microwave radiometry', *J. Geophys. Res. Atmos.*, vol. 100, no. D11, 1995.

[19] J. Chen, Y. Chen, and J. Yang, 'Ship detection using polarization cross-entropy', *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 723–727, 2009.

[20] E. Chuvieco and A. Huete, *Fundamentals of satellite remote sensing*. Florida, USA: CRC Press, 2009.

[21] J. R. Jensen, *Remote sensing of the environment: an earth resource perspective*, 2nd ed. Harlow, England: Pearson Education India, 2009.

[22] USGS, 'What are the band designations for the Landsat satellites?', *USGS*, 2014. [Online]. Available: https://www.usgs.gov/faqs/what-are-band-designations-landsat-satellites?qt-news_science_products=0#qt-news_science_products. [Accessed: 23-Dec-2020].

[23] European Spatial Agency (ESA), 'MSI Instrument – Sentinel-2 MSI Technical Guide – Sentinel Online', *Sentinel Online*, 2019. [Online]. Available: https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument. [Accessed: 23-Dec-2020].

[24] S. Mishra, D. R. Mishra, and W. M. Schluchter, 'A novel algorithm for predicting phycocyanin concentrations in cyanobacteria: A proximal hyperspectral remote sensing

approach', *Remote Sens.*, vol. 1, no. 4, pp. 758–775, 2009.

[25] LAADS DAAC, 'Moderate Resolution Imaging Spectroradiometer (MODIS) - LAADS DAAC', *NASA*, 2020. [Online]. Available: https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/modis/. [Accessed: 23-Dec-2020].

[26] S. C. Liew, 'Interpreting Optical Remote Sensing Images', *Principles of Remote Sensing*, vol. 2. Centre for Remote Imaging, Sensing and Processing, CRISP, pp. 2–6, 2001.

[27] F. T. Ulaby, R. K. Moore, and A. K. Fung, *Microwave remote sensing: active and passive. Volume I: microwave remote sensing fundamentals and radiometry.*, 1st ed. Norwood, MA: Artech House, 1981.

[28] S. Riegger, W. Wiesbeck, and A. J. Sieber, 'On the origin of cross-polarization in remote sensing', in *International Geoscience and Remote Sensing Symposium,* 1987, pp. 577–580.

[29] V. Sambasiva Rao, V. V. Srinivasan, and S. Pal, 'Generation of dual beams from spherical phased array antenna', *Electron. Lett.*, vol. 45, no. 9, pp. 441–442, 2009.

[30] F. Meyer, 'Spaceborne Synthetic Aperture Radar: Principles, data access, and basic processing techniques', 2019.

[31] P. K. Jain, S. Pal, and V. M. Pandharipande, 'Real-time quality analysis of digital remote sensing satellite signals', *RF Des.*, vol. 26, no. 3, pp. 14–21, 2003.

[32] 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.

[33] M. Saraswat and K. V. Arya, 'Feature selection and classification of leukocytes using random forest', *Med. Biol. Eng. Comput.*, vol. 52, no. 12, pp. 1041–1052, 2014.

[34] Y. Li, H. Zhang, X. Xue, Y. Jiang, and Q. Shen, 'Deep learning for remote sensing image classification: A survey', *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 6, Nov. 2018.

[35] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, 'Deep learning', *Nature*, vol. 321, pp. 436–444, 2015.

[36] C. Yonezawa, M. Watanabe, and G. Saito, 'Polarimetric decomposition analysis of ALOS PALSAR observation data before and after a landslide event', *Remote Sens.*, vol. 4, no. 8, pp. 2314–2328, 2012.

[37] Y. Cui *et al.*, 'On complete model-based decomposition of polarimetric SAR coherency matrix data', *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 1991–2001, 2013.

[38] S. W. Chen, X. S. Wang, S. P. Xiao, and M. Sato, 'Generalized model-based

decomposition for polarimetric SAR data', *IET Conf. Publ.*, vol. 2013, no. 617 CP, pp. 2732–2739, 2013.

[39] G. Singh *et al.*, 'Seven-component scattering power decomposition of PolSAR coherency matrix', *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8371–8382, 2019.

[40] G. Singh, G. Venkataraman, Y. Yamaguchi, and S. E. Park, 'Capability assessment of fully polarimetric alos-palsar data for discriminating wet snow from other scattering types in mountainous regions', *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1177–1196, 2014.

[41] W. An, C. Xie, X. Yuan, Y. Cui, and J. Yang, 'Four-component decomposition of polarimetric SAR images with deorientation', *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 1090–1094, 2011.

[42] D. Ratha, E. Pottier, A. Bhattacharya, and A. C. Frery, 'A PolSAR scattering power factorization framework and novel roll-invariant parameters based unsupervised classification scheme using a geodesic distance', *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3509–3525, 2019.

[43] S. Gopal, C. E. Woodcock, and A. H. Strahler, 'Fuzzy neural network classification of global land cover from a 1° AVHRR data set', *Remote Sens. Environ.*, vol. 67, no. 2, pp. 230–243, 1999.

[44] N. Joshi *et al.*, 'A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring', *Remote Sens.*, vol. 8, no. 1, pp. 1–23, 2016.

[45] D. J. Lary, A. H. Alavi, A. H. Gandomi, and A. L. Walker, 'Machine learning in geosciences and remote sensing', *Geosci. Front.*, vol. 7, no. 1, pp. 3–10, 2016.

[46] P. Feng, B. Wang, D. L. Liu, and Q. Yu, 'Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia', *Agric. Syst.*, vol. 173, pp. 303–316, 2019.

[47] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, 'A tutorial on synthetic aperture radar', *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, 2013.

[48] J. Yang, G. Dong, Y. Peng, Y. Yamaguchi, and H. Yamada, 'Generalized optimization of polarimetric contrast enhancement', *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 3, pp. 171–174, 2004.

[49] G. Zhou, Y. Cu, Y. Chen, J. Yang, H. Rishvand, and Y. Yamaguchi, 'Linear feature detection in Polarimetric SAR images', *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1453–1463, 2011.

[50] C. He, S. Li, Z. Liao, and M. Liao, 'Texture classification of PolSAR data based on sparse

coding of wavelet polarization textons', *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 8, pp. 4576–4590, 2013.

[51] J. Chen, Y. Chen, W. An, Y. Cui, and J. Yang, 'Nonlocal filtering for polarimetric SAR data: A pretest approach', *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1744–1754, 2011.

[52] C. Pohl and J. L. Van Genderen, 'Multisensor image fusion in remote sensing: Concepts, methods and applications', *Int. J. Remote Sens.*, vol. 19, no. 5, pp. 823–854, 1998.

[53] L. Zhu and R. Tateishi, 'Fusion of multisensor multitemporal satellite data for land cover mapping', *Int. J. Remote Sens.*, vol. 27, no. 5, pp. 903–918, 2006.

[54] J. Dong, D. Zhuang, Y. Huang, and J. Fu, 'Advances in multi-sensor data fusion: Algorithms and applications', *Sensors*, vol. 9, no. 10, pp. 7771–7784, 2009.

[55] J. Zhang, 'Multi-source remote sensing data fusion: Status and trends', *Int. J. Image Data Fusion*, vol. 1, no. 1, pp. 5–24, Mar. 2010.

[56] D. Mishra, S. Narumalani, M. Lawson, and D. Rundquist, 'Bathymetric mapping using IKONOS multispectral data', *GIScience Remote Sens.*, vol. 41, no. 4, pp. 301–321, 2004.

[57] C. Pohl, *Geometric aspects of multisensor image fusion for topographic map updating in the humid tropics*, vol. 39. 1996.

[58] X. X. Zhu *et al.*, 'Deep learning in remote sensing: A comprehensive review and list of resources', *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8-36, 2017.

[59] D. Bargiel, S. Herrmann, D. Bargiel, and S. Herrmann, 'Multi-temporal land-cover classification of agricultural areas in two european regions with high resolution spotlight TerraSAR-X data', *Remote Sens.*, vol. 3, no. 5, pp. 859–877, Apr. 2011.

[60] L. Gomez-Chova, D. Tuia, G. Moser, and G. Camps-Valls, 'Multimodal classification of remote sensing images: A review and future directions', *Proc. IEEE*, vol. 103, no. 9, pp. 1560–1584, 2015.

[61] C. Gómez, J. C. White, and M. A. Wulder, 'Optical remotely sensed time series data for land cover classification: A review', *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 55–72, Jun. 2016.

[62] J. Gong, H. Sui, G. Ma, and Q. Zhou, 'A review of multi-temporal remote sensing data change detection algorithms', *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 37, pp. 757–762, 2008.

[63] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, 'A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges', *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, 2019.

[64] B. P. Page, A. Kumar, and D. R. Mishra, 'A novel cross-satellite based assessment of the

spatio-temporal development of a cyanobacterial harmful algal bloom', *Int. J. Appl. Earth Obs. Geoinf.*, vol. 66, no. 4, pp. 69–81, 2018.

[65] C. L. M. de O. Santos *et al.*, 'Classification of crops, pastures, and tree plantations along the season with multi-sensor image time series in a subtropical agricultural region', *Remote Sens.*, vol. 11, no. 3, pp. 334–360, 2019.

[66] H. Skriver, 'Crop classification by multitemporal C- and L-band single- and dual-polarization and fully polarimetric SAR', *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2138–2149, 2012.

[67] P. Villa, D. Stroppiana, G. Fontanelli, R. Azar, and P. A. Brivio, 'In-season mapping of crop type with optical and X-band SAR data: A classification tree approach using synoptic seasonal features', *Remote Sens.*, vol. 7, no. 10, pp. 12859–12886, 2015.

[68] H. McNairn and J. Shang, 'A Review of Multitemporal Synthetic Aperture Radar (SAR) for Crop Monitoring', in *Multitemporal Remote Sensing: Methods and Applications*, vol. 20, Y. Ban, Ed. Springer International Publishing, 2016, pp. 317–340.

[69] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, 'Deep learning in remote sensing applications: A meta-analysis and review', *ISPRS J. Photogramm. Remote Sens.*, vol. 152, no. 6, pp. 166–177, 2019.

[70] A. Kamilaris and F. X. Prenafeta-Boldú, 'Deep learning in agriculture: A survey', *Comput. Electron. Agric.*, vol. 147, no. 4, pp. 70–90, 2018.

[71] 'A.I. technical - Machine Learning vs. Deep Learning'. [Online]. Available: https://lawtomated.com/a-i-technical-machine-vs-deep-learning/. [Accessed: 24-Dec-2020].

[72] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, 'Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data', *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, 2017.

[73] C. Pelletier, G. I. Webb, and F. Petitjean, 'Temporal convolutional neural network for the classification of satellite image time series', *Remote Sens.*, vol. 11, no. 5, 2019.

[74] M. Rubwurm, M. Korner, M. Rußwurm, and M. Körner, 'Temporal Vegetation Modelling Using Long Short-Term Memory Networks for Crop Identification from Medium-Resolution Multi-spectral Satellite Images', in *Computer Vision and Pattern Recognition Workshops*, 2017, pp. 1496–1504.

[75] H. Parikh, S. Patel, and V. Patel, 'Classification of SAR and PolSAR images using deep learning: a review', *Int. J. Image Data Fusion*, vol. 11, no. 1, pp. 1–32, 2020.

[76] Q. Lv, Y. Dou, X. Niu, J. Xu, J. Xu, and F. Xia, 'Urban land use and land cover classification using remotely sensed sar data through deep belief networks', *J. Sensors*,

vol. 2015, 2015.

[77] S. Chen and H. Wang, 'SAR target recognition based on deep learning', in *International Conference on Data Science and Advanced Analytics*, 2014, pp. 541–547.

[78] L. Wang, K. A. Scott, L. Xu, and D. A. Clausi, 'Sea Ice Concentration Estimation during Melt from Dual-Pol SAR Scenes Using Deep Convolutional Neural Networks: A Case Study', *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4524–4533, 2016.

[79] P. Wang, H. Zhang, and V. M. Patel, 'SAR image despeckling using a convolutional neural network', *arXiv*, vol. 24, no. 12, pp. 1763–1767, 2017.

[80] N. Anantrasirichai, J. Biggs, F. Albino, P. Hill, and D. Bull, 'Application of Machine Learning to Classification of Volcanic Deformation in Routinely Generated InSAR Data', *J. Geophys. Res. Solid Earth*, vol. 123, no. 8, pp. 6592–6606, 2018.

[81] H. Xie, S. Wang, K. Liu, S. Lin, and B. Hou, 'Multilayer Feature Learning for Polarimetric Synthetic Radar Data Classification', in *Geoscience and Remote Sensing Symposium*, 2014, pp. 2818–2821.

[82] J. Geng, J. Fan, H. Wang, X. Ma, B. Li, and F. Chen, 'High-Resolution SAR Image Classification via Deep Convolutional Autoencoders', *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2351–2355, 2015.

[83] F. Qin, J. Guo, and W. Sun, 'Object-oriented ensemble classification for polarimetric SAR Imagery using restricted Boltzmann machines', *Remote Sens. Lett.*, vol. 8, no. 3, pp. 204–213, 2017.

[84] L. Zhang, W. Ma, and D. Zhang, 'Stacked Sparse Autoencoder in PolSAR Data Classification Using Local Spatial Information', *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1359–1363, 2016.

[85] Z. Zhao, L. Jiao, J. Zhao, J. Gu, and J. Zhao, 'Discriminant deep belief network for high-resolution SAR image classification', *Pattern Recognit.*, vol. 61, pp. 686–701, 2017.

[86] F. Gao, T. Huang, J. Wang, J. Sun, A. Hussain, and E. Yang, 'Dual-branch deep convolution neural network for polarimetric SAR image classification', *Appl. Sci.*, vol. 7, no. 5, 2017.

[87] Y. Zhou, H. Wang, F. Xu, and Y. Q. Jin, 'Polarimetric SAR Image Classification Using Deep Convolutional Neural Networks', *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1935–1939, 2016.

[88] H. Dong, L. Zhang, and B. Zou, 'PolSAR image classification with lightweight 3D convolutional networks', *Remote Sens.*, vol. 12, no. 3, 2020.

[89] Y. Wang, C. He, X. Liu, and M. Liao, 'A hierarchical fully convolutional network integrated with sparse and low-rank subspace representations for PolSAR imagery

classification', *Remote Sens.*, vol. 10, no. 2, 2018.

[90] S. W. Chen and C. S. Tao, 'PolSAR image classification using polarimetric-feature-driven deep convolutional neural network', *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 627–631, 2018.

[91] C. He, M. Tu, D. Xiong, and M. Liao, 'Nonlinear manifold learning integrated with fully convolutional networks for PolSAR image classification', *Remote Sens.*, vol. 12, no. 4, 2020.

[92] Z. Zhang, H. Wang, F. Xu, and Y. Q. Jin, 'Complex-valued convolutional neural network and its application in polarimetric SAR image classification', *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7177–7188, 2017.

[93] A. G. Mullissa, C. Persello, and A. Stein, 'Polsarnet: A deep fully convolutional network for polarimetric sar image classification', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 12, pp. 5300–5309, 2019.

[94] L. Li, L. Ma, L. Jiao, F. Liu, Q. Sun, and J. Zhao, 'Complex contourlet-CNN for polarimetric SAR image classification', *Pattern Recognit.*, vol. 100, no. 4, p. 107110, 2020.

[95] W. Xie, G. Ma, F. Zhao, H. Liu, and L. Zhang, 'PolSAR image classification via a novel semi-supervised recurrent complex-valued convolution neural network', *Neurocomputing*, vol. 388, no. 19, pp. 255–268, 2020.

[96] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, 'Deep feature extraction and classification of hyperspectral images based on convolutional neural networks', *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, 2016.

[97] Y. Chen, X. Zhao, S. Member, X. Jia, and S. Member, 'Spectral – spatial classification of hyperspectral data based on deep belief network', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, 2015.

[98] Y. Li, H. Zhang, and Q. Shen, 'Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network', *Remote Sens.*, vol. 9, no. 1, p. 67, 2017.

[99] S. N. Chaudhri, N. S. Rajput, K. P. Singh, and D. Singh, 'Maximum membership fraction based pure pixel assessment approach for hyperspectral data analysis using deep learning', in *International Geoscience and Remote Sensing Symposium*, 2019, pp. 5820–5823.

[100] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, 'Deep convolutional neural networks for hyperspectral image classification', *J. Sensors*, vol. 2015, pp. 1–12, 2015.

[101] S. Mei, J. Ji, Q. Bi, J. Hou, Q. Du, and W. Li, 'Integrating spectral and spatial information into deep convolutional Neural Networks for hyperspectral classification', in *International Geoscience and Remote Sensing Symposium*, 2016, pp. 5067–5070.

[102] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, 'Deep supervised learning for hyperspectral data classification through convolutional neural networks', in *International Geoscience and Remote Sensing Symposium*, 2015, pp. 4959–4962.

[103] J. Yue, S. Mao, and M. Li, 'A deep learning framework for hyperspectral image classification using spatial pyramid pooling', *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 9, pp. 875–884, 2016.

[104] W. Zhao, Z. Guo, J. Yue, X. Zhang, and L. Luo, 'On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery', *Int. J. Remote Sens.*, vol. 36, no. 13, pp. 3368–3379, 2015.

[105] E. Aptoula, M. C. Ozdemir, and B. Yanikoglu, 'Deep learning with attribute profiles for hyperspectral image classification', *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1970–1974, 2016.

[106] H. Liang and Q. Li, 'Hyperspectral imagery classification using sparse representations of convolutional neural network features', *Remote Sens.*, vol. 8, no. 2, p. 99, 2016.

[107] Y. Li, W. Xie, and H. Li, 'Hyperspectral image reconstruction by deep convolutional neural network for classification', *Pattern Recognit.*, vol. 63, no. 3, pp. 371–383, 2017.

[108] H. Zhang, Y. Li, Y. Zhang, and Q. Shen, 'Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network', *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 438–447, 2017.

[109] F. Hu, G. S. Xia, Z. Wang, X. Huang, L. Zhang, and H. Sun, 'Unsupervised feature learning via spectral clustering of multidimensional patches for remotely sensed scene classification', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 5, pp. 2015–2030, May 2015.

[110] F. Hu, G. S. Xia, J. Hu, Y. Zhong, and K. Xu, 'Fast binary coding for the scene classification of high-resolution remote sensing imagery', *Remote Sens.*, vol. 8, no. 7, p. 555, 2016.

[111] G. S. Xia *et al.*, 'AID: A benchmark data set for performance evaluation of aerial scene classification', *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, 2017.

[112] A. Sharma, X. Liu, X. Yang, and D. Shi, 'A patch-based convolutional neural network for remote sensing image classification', *Neural Networks*, vol. 95, no. 11, pp. 19–28, 2017.

[113] L. Loncan *et al.*, 'Hyperspectral pansharpening: A review', *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, 2015.

[114] D. Fasbender, D. Tuia, P. Bogaert, and M. Kanevski, 'Support-based implementation of bayesian data fusion for spatial enhancement: applications to ASTER thermal images', *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 598–602, 2008.

[115] J. Zhong, B. Yang, G. Huang, F. Zhong, and Z. Chen, 'Remote sensing image fusion with convolutional neural network', *Sens. Imaging*, vol. 17, no. 1, p. 10, 2016.

[116] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, 'Pansharpening by convolutional neural networks', *Remote Sens.*, vol. 8, no. 7, p. 594, 2016.

[117] Y. Yuan, X. Zheng, and X. Lu, 'Hyperspectral image superresolution by transfer learning', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 5, pp. 1963–1974, 2017.

[118] D. Tuia, C. Persello, and L. Bruzzone, 'Recent advances in domain adaptation for the classification of remote sensing data', *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 41–57, 2016.

[119] M. Volpi and D. Tuia, 'Dense semantic labeling of subdecimeter resolution images with convolutional neural networks', *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 881–893, 2017.

[120] J. Geng, H. Wang, J. Fan, and X. Ma, 'Deep supervised and contractive neural network for sar image classification', *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2442–2459, 2017.

[121] D. Tuia, R. Flamary, and N. Courty, 'Multiclass feature learning for hyperspectral image classification: sparse and hierarchical solutions', *ISPRS J. Photogramm. Remote Sens.*, vol. 105, no. 7, pp. 272–285, 2015.

[122] S. Paisitkriangkrai, J. Sherrah, P. Janney, and A. van den Hengel, 'Semantic labeling of aerial and satellite imagery', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 9, no. 7, pp. 2868–2881, 2016.

[123] D. Marcos, R. Hamid, and D. Tuia, 'Geospatial correspondences for multimodal registration', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5091–5100.

[124] N. Di Mauro, A. Vergari, T. M. A. Basile, F. G. Ventola, and F. Esposito, 'End-to-end learning of deep spatio-temporal representations for satellite image time series classification', *CEUR Workshop Proc.*, vol. 1972, 2017.

[125] Y. Zheng, Q. Liu, E. Chen, Y. Ge, and J. L. Zhao, 'Time series classification using multi-channels deep convolutional neural networks', in *International Conference on Web-Age Information Management*, 2014, vol. 8485, pp. 298–310.

[126] L. Zhong, L. Hu, and H. Zhou, 'Deep learning based multi-temporal crop classification', *Remote Sens. Environ.*, vol. 221, no. 11, pp. 430–443, 2019.

[127] S. Ji, C. Zhang, A. Xu, Y. Shi, and Y. Duan, '3D convolutional neural networks for crop classification with multi-temporal remote sensing images', *Remote Sens.*, vol. 10, no. 1, p. 75, 2018.

[128] Z. Li, G. Chen, and T. Zhang, 'A CNN-transformer hybrid approach for crop classification using multitemporal multisensor images', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, no. 1, pp. 847–858, 2020.

[129] H. Lyu, H. Lu, and L. Mou, 'Learning a transferable change rule from a recurrent neural network for land cover change detection', *Remote Sens.*, vol. 8, no. 6, p. 506, 2016.

[130] D. Ienco, R. Gaetano, C. Dupaquier, and P. Maurel, 'Land cover classification via multioral spatial data by recurrent neural networks', *arXiv*, vol. 14, no. 10, pp. 1685–1689, 2017.

[131] E. Ndikumana, D. H. T. Minh, N. Baghdadi, D. Courault, and L. Hossard, 'Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue, France', *Remote Sens.*, vol. 10, no. 8, p. 1217, 2018.

[132] D. Ho Tong Minh *et al.*, 'Deep recurrent neural networks for winter vegetation quality mapping via multitemporal SAR sentinel-1', *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 465–468, 2018.

[133] Z. Sun, L. Di, and H. Fang, 'Using long short-term memory recurrent neural network in land cover classification on landsat and cropland data layer time series', *Int. J. Remote Sens.*, vol. 40, no. 2, pp. 593–614, 2019.

[134] A. Sharma, X. Liu, and X. Yang, 'Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks', *Neural Networks*, vol. 105, no. 9, pp. 346–355, 2018.

[135] P. Benedetti, D. Ienco, R. Gaetano, K. Ose, R. G. Pensa, and S. Dupuy, ''M3 fusion: A deep learning architecture for multiscale multimodal multitemporal satellite data fusion', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 12, pp. 4939–4949, 2018.

[136] K. K. Gadiraju, B. Ramachandra, Z. Chen, and R. R. Vatsavai, 'Multimodal deep learning based crop classification using multispectral and multitemporal satellite imagery', in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2020, pp. 3234–3242.

[137] L. Mou, L. Bruzzone, and X. X. Zhu, 'Learning spectral-spatialoral features via a recurrent convolutional neural network for change detection in multispectral imagery', *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 924–935, 2019.

[138] J. Sen Lee and E. Pottier, *Polarimetric radar imaging: From basics to applications*. Boca Raton, FL, USA: CRC Press, 2017.

[139] S. R. Cloude and E. Pottier, 'An entropy based classification scheme for land applications of polarimetric SAR', *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 68–78, 1997.

[140] S. R. Cloude and P. Eric, 'A review of target decomposition theorems in radar

polarimetry', *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 498–518, 1996.

[141] P. Mishra, A. Garg, and D. Singh, 'Critical analysis of model-based incoherent polarimetric decomposition methods and investigation of deorientation effect', *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 4868–4877, 2017.

[142] A. Kumar, R. K. Panigrahi, and A. Das, 'Three-component decomposition technique for hybrid-pol SAR data', *IET Radar, Sonar Navig.*, vol. 10, no. 9, pp. 1569–1574, 2016.

[143] H. Wang, Z. Zhou, J. Turnbull, Q. Song, and F. Qi, 'Pol-SAR classification based on generalized polar decomposition of mueller matrix', *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 4, pp. 565–569, 2016.

[144] J. Wei, Z. Zhao, X. Yu, and L. Lu, 'A multi-component decomposition method for polarimetric SAR data', *Chinese J. Electron.*, vol. 26, no. 1, pp. 205–210, 2017.

[145] V. Turkar and Y. S. Rao, 'Effect of different target decomposition techniques on classificatio accuracy for polarimetric SAR data', in *Technology Systems and Management*, 2011, pp. 138–145.

[146] S. De, L. Bruzzone, A. Bhattacharya, F. Bovolo, and S. Chaudhuri, 'A novel technique based on deep learning and a synthetic target database for classification of urban areas in PolSAR data', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 1, pp. 154–170, 2018.

[147] D. Ratha, D. Mandal, V. Kumar, H. Mcnairn, A. Bhattacharya, and A. C. Frery, 'A generalized volume scattering model-based vegetation index from polarimetric SAR data', *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 11, pp. 1791–1795, 2019.

[148] H. Bi, F. Xu, Z. Wei, Y. Xue, and Z. Xu, 'An active deep learning approach for minimally supervised PolSAR image classification', *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9378–9395, 2019.

[149] M. Jafari, Y. Maghsoudi, and M. J. V. Zoej, 'A new method for land cover characterization and classification of polarimetric SAR data using polarimetric signatures', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 7, pp. 3595–3607, 2015.

[150] M. Bielecka, S. Porzycka-Strzelczyk, and J. Strzelczyk, 'SAR images analysis based on polarimetric signatures', *Appl. Soft Comput. J.*, vol. 23, pp. 259–269, 2014.

[151] F. Sadjadi, 'Improved target classification using optimum polarimetric SAR signatures', *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 1, pp. 38–49, 2002.

[152] K. Arai and J. Wang, 'Polarimetric SAR image classification with the maximum curvature of the trajectory in the eigen space converted from the polarization signature', *Adv. Sp. Res.*, vol. 39, no. 1, pp. 149–154, 2007.

[153] J. Strzelczyk and S. Porzycka-Strzelczyk, 'Identification of coherent scatterers in SAR

images based on the analysis of polarimetric signatures', *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 4, pp. 783–787, 2014.

[154] J. J. van Zyl, H. A. Zebker, and C. Elachi, 'Imaging radar polarization signatures: Theory and observation', *Radio Sci.*, vol. 22, no. 4, pp. 529–543, 1987.

[155] E. Kennaugh and R. Sloan, 'Effects of type of polarization on echo characteristics', Columbus, Ohio, 1952.

[156] Jong-Sen Lee, Mitchell R. Grunes, and Wolfgang-Martin Boerner;, 'Polarimetric property preservation in SAR speckle filtering', in *Proc. SPIE 3120, Wideband Interferometric Sensing and Imaging Polarimetry*, 1997, pp. 1–7.

[157] Ian Goodfellow, Y. Bengio, and A. Courville, 'Convolutional Neural Networks', in *Deep learning*, 1st ed., vol. 13, no. 1, Cambridge: Cambridge, MIT Press, 2017, pp. 330–372.

[158] D. Ciresan, U. Meier, and J. Schmidhuber, 'Multi-column deep neural networks for image classification', in *IEEE conference on Computer Vision and Pattern Recognition*, 2012, pp. 3642–3649.

[159] K. Fukushima and S. Miyake, 'Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition', in *Competition and cooperation in neural nets*, Berlin, Heidelberg.: Springer, 1982, pp. 267–285.

[160] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, 'Gradient-based learning applied to document recognition', *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[161] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, vol. 1, no. 2. Massachusetts, 2017.

[162] S. Albelwi and A. Mahmood, 'A framework for designing the architectures of deep Convolutional Neural Networks', *Entropy*, vol. 19, no. 6, pp. 242-, 2017.

[163] Y. L. Boureau, J. Ponce, J. P. Fr, and Y. Lecun, 'A theoretical analysis of feature pooling in visual recognition', in *international Conference on Machine Learning*, 2010, pp. 111–118.

[164] A. B. Pour and M. Hashim, 'Structural geology mapping using PALSAR data in the Bau gold mining district, Sarawak, Malaysia', *Adv. Sp. Res.*, vol. 54, no. 4, pp. 644–654, 2014.

[165] P. E. Yastika, N. Shimizu, and H. Z. Abidin, 'Monitoring of long-term land subsidence from 2003 to 2017 in coastal area of Semarang, Indonesia by SBAS DInSAR analyses using Envisat-ASAR, ALOS-PALSAR, and Sentinel-1A SAR data', *Adv. Sp. Res.*, vol. 63, no. 5, pp. 1719–1736, 2019.

[166] A. Jain and D. Singh, 'An optimal selection of probability distribution functions for unsupervised land cover classification of PALSAR-2 data', *Adv. Sp. Res.*, vol. 63, no. 2, pp. 813–825, 2019.

[167] V. Mittal, D. Singh, and L. M. Saini, 'A critical analysis of em based fusion of different polarization data for effect on land cover classification', *Adv. Sp. Res.*, vol. 56, no. 6, pp. 1094–1105, 2015.

[168] P. Mishra, D. Singh, and Y. Yamaguchi, 'Land cover classification of PALSAR images by knowledge based decision tree classifier and supervised classifiers based on SAR observables', *Prog. Electromagn. Res.*, vol. 30, no. 30, pp. 47–70, 2011.

[169] E. Nezry and F. Yakam-simen, 'On the preservation of polarimetric signatures and polarimetric texture signatures by fully polarimetric map filters', in *Geoscience and Remote Sensing Symposium*, 1999, pp. 1555–1557.

[170] P. Mishra and D. Singh, 'A statistical-measure-based adaptive land cover classification algorithm by efficient utilization of polarimetric SAR observables', *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2889–2900, 2014.

[171] Y. G. Cao, L. J. Yan, and Z. Z. Zheng, 'Extraction of information on geology hazard from multi-polarization SAR images', in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2008, vol. 37, no. B4, pp. 1529–1532.

[172] A. K. Maurya, 'An approach to use polarimetric signature for land cover classification', in *National Conference on Recent Advances in Electronics & Computer Engineering*, 2015, pp. 248–253.

[173] N. Phruksahiran and M. Chandra, 'Backscatter power measurement of canonical targets under GBSAR environment', in *European Conference on Synthetic Aperture Radar*, 2016, pp. 290–292.

[174] J.-S. Lee and E. Pottier, 'Electromagnetic Vector Scattering Operators', in *Polarimetric Radar Imaging: From Basics to Applications*, 1st ed., Brian J. Thompson, Ed. New York: CRC Press, 2009, pp. 92–98.

[175] I. G. Cumming and J. J. Van Zyl, 'Feature utility in polarimetric radar image classification', in *Geoscience and Remote Sensing Symposium*, 1989, pp. 1841–1846.

[176] A. Bhattacharya, G. Singh, S. Manickam, and Y. Yamaguchi, 'An adaptive general four-component scattering power decomposition with unitary transformation of coherency matrix (AG4U)', *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 10, pp. 2110–2114, 2015.

[177] J. Sen Lee, T. L. Ainsworth, and Y. Wang, 'Generalized polarimetric model-based decompositions using incoherent scattering models', *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2474–2491, 2014.

[178] E. Pottier and J.-S. Lee., 'Application of the H/A/alpha polarimetric decomposition theorem for unsupervised classification of fully polarimetric SAR data based on the wishart distribution', in *SAR workshop: CEOS Committee on Earth Observation Satellites*,

1997, pp. 132–143.

[179] G. S. Phartiyal, K. Kumar, D. Singh, and K. P. Singh, 'Optimal use of polarimetric ssignature on PALSAR -2 data for land cover classification', in *Geoscience and Remote Sensing Symposium*, 2017, pp. 4558–4561.

[180] D. E. Goldberg and Jon Richardson, 'Genetic algorithms with sharing for multimodal function optimization', in *International Conference on Genetic Algorithms on Genetic algorithms and their application*, 1987, pp. 41–49.

[181] J. C. Bansal, H. Sharma, K. V. Arya, K. Deep, and M. Pant, 'Self-adaptive artificial bee colony', *Optimization*, vol. 63, no. 10, pp. 1513–1532, 2014.

[182] A. Garg and D. Singh, 'Development of an efficient contextual algorithm for discrimination of tall vegetation and urban for PALSAR data', *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3413–3420, 2018.

[183] L. Giustarini, R. Hostache, P. Matgen, G. J. P. Schumann, P. D. Bates, and D. C. Mason, 'A change detection approach to flood mapping in Urban areas using TerraSAR-X', *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2417–2430, 2013.

[184] R. K. Sharma, B. S. Kumar, N. M. Desai, and V. R. Gujraty, 'SAR for disaster management', *IEEE Aerospace and Electronic Systems*, vol. 23, no. 6, pp. 4–9, 2008.

[185] S. Martinis, J. Kersten, and A. Twele, 'A fully automated TerraSAR-X based flood service', *ISPRS J. Photogramm. Remote Sens.*, vol. 104, pp. 203–212, 2015.

[186] J. Jacinth Jennifer, S. Saravanan, and D. Abijith, 'Integration of SAR and multi-spectral imagery in flood inundation mapping–a case study on Kerala floods 2018', *ISH J. Hydraul. Eng.*, vol. 00, pp. 1–11, 2020.

[187] B. Pradhan, M. S. Tehrany, and M. N. Jebur, 'A new semiautomated detection mapping of flood extent from TerraSAR-X satellite image using rule-based classification and taguchi optimization techniques', *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4331–4342, 2016.

[188] E. Schnebele, G. Cervone, S. Kumar, and N. Waters, 'Real time estimation of the calgary floods using limited remote sensing data', *Water*, vol. 6, no. 2, pp. 381–398, 2014.

[189] Dasgupta A., Grimaldi S., Ramakrishnan R., and Walker J. P., 'Optimized GLCM ¬ based texture features for improved SAR ¬ based flood mapping', in *Geoscience and Remote Sensing Symposium*, 2017, pp. 3258–3261.

[190] Y. Yamaguchi, *Polarimetric SAR Imaging*, 1st ed. Bota Racon, Florida: CRC Press, 2020.

[191] S. Plank, M. Jüssi, S. Martinis, and A. Twele, 'Mapping of flooded vegetation by means of polarimetric sentinel-1 and ALOS-2/PALSAR-2 imagery', *Int. J. Remote Sens.*, vol. 38, no. 13, pp. 3831–3850, 2017.

[192] G. D. Silvana and G. Laura, 'Adaptive sar image processing techniques to support flood monitoring from earth observation data', in *Flood Monitoring through Remote Sensing*, 1st ed., A. Refice, A. D'Addabbo, and D. Capolongo, Eds. Bari, Italy: Springer, 2018.

[193] A. K. Maurya, G. S. Phartiyal, and D. Singh, 'A critical analysis of polarimetric signatures on PALSAR 2 data for land cover classification', in *IEEE International Conference on Industrial and Information Systems*, 2016, vol. 11, pp. 920–924.

[194] O. . De Carvalho and P. R. Meneses, 'Spectral correlation mapper (SCM): an improvement on the spectral angle mapper (SAM)', in *Summaries of the 9th JPL Airborne Earth Science Workshop*, 2000, vol. 9.

[195] J. A. Palmason, J. A. Benediktsson, and J. R. Sveinsson, 'Classification of hyperspectral data from urban areas based on extended morphological profile', *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 63–69, 2005.

[196] J. C. Gower, 'Properties of Euclidean and non-Euclidean distance matrices', *Linear Algebra Appl.*, vol. 67, no. C, pp. 81–97, 1985.

[197] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Prentice Hall, Pearson, 2008.

[198] S. N. Chaudhri, N. S. Rajput, K. P. Singh, and D. Singh, 'Different Modality Based Remote Sensing Data Fusion Approach for Efficient classification of agriculture and urban subclasses', in *International Geoscience and Remote Sensing Symposium*, 2019, pp. 5710–5713.

[199] S. Lawrence and C. L. Giles, 'Overfitting and neural networks: conjugate gradient and backpropagation', in *International Joint Conference on Neural Networks Neural Computing: New Challenges and Perspectives for the New Millennium*, 2000, pp. 114–119.

[200] B. Huang, B. Zhao, and Y. Song, 'Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery', *Remote Sens. Environ.*, vol. 214, pp. 73–86, 2018.

[201] S. S. Rajput and K. V. Arya, 'CNN classifier based low-resolution Face recognition algorithm', in *International Conference on Emerging Frontiers in Electrical and Electronic Technologies*, 2020.

[202] R. S. Bhadoria, J. R. Pinto, J. S. Cardoso, and A. Lourenço, 'Deep neural networks for biometric identification based on non-intrusive ecg acquisitions', in *The Biometric Computing: Recognition and Registration*, 1st ed., K. V. Arya and R. S. Bhadoria, Eds. CRC Press, 2020, pp. 217–234.

[203] J. Feng *et al.*, 'CNN-based multilayer spatial-spectral feature fusion and sample

augmentation with local and nonlocal constraints for hyperspectral image classification', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 4, pp. 1299–1313, 2019.

[204] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, 'Deep learning-based classification of hyperspectral data', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, 2014.

[205] M. He, X. Li, Y. Zhang, J. Zhang, and W. Wang, 'Hyperspectral image classification based on deep stacking network', in *International Geoscience and Remote Sensing Symposium*, 2016, pp. 3286–3289.

[206] S. Yu, S. Jia, and C. Xu, 'Convolutional neural networks for hyperspectral image classification', *Neurocomputing*, vol. 219, no. 1, pp. 88–98, 2017.

[207] J. Yue, W. Zhao, S. Mao, and H. Liu, 'Spectral-spatial classification of hyperspectral images using deep convolutional neural networks', *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 6, pp. 468–477, 2015.

[208] W. Zhao and S. Du, 'Spectral-spatial feature extraction for hyperspectral image classification: a dimension reduction and deep learning approach', *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, 2016.

[209] A. R. Gillespie, A. B. Kahle, and R. E. Walker, 'Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques', *Remote Sens. Environ.*, vol. 22, no. 3, pp. 343–365, 1987.

[210] U. K. Rajput, S. K. Ghosh, and A. Kumar, 'Multisensor fusion of satellite images for urban information extraction using pseudo-Wigner distribution', *J. Appl. Remote Sens.*, vol. 8, no. 1, p. 083668, 2014.

[211] D. P. Kingma and J. L. Ba, 'Adam: A method for stochastic optimization', *arXiv*, vol. 1412, no. 6980, Dec. 2014.

[212] S. S. Heydari and G. Mountrakis, 'Meta-analysis of deep neural networks in remote sensing: A comparative study of mono-temporal classification to support vector machines', *ISPRS J. Photogramm. Remote Sens.*, vol. 152, no. 6, pp. 192–210, 2019.

[213] T. Garnett and C. Godfray, 'Sustainable intensification in agriculture: Navigating a course through competing food system priorities', UK, 2012.

[214] T. W. Bank, The International Bank for Reconstruction and Development, 'Agricultural Innovation Systems An Investment Sourcebook', Washington DC, 2011.

[215] Inter-Agency and Expert Group on SDG Indicators (IAEG-SDGs), 'Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development', new york, 2017.

[216] S. Godbole, A. Kulkarni, and N. Despande, 'Remote Sensing: Applications in

Agriculture', in *National Conference on Emerging Trends in Information Technology*, 2007, p. 348.

[217] V. I. Myers, 'Remote sensing applications in agriculture', in *Manual of remote sensing-second edition: volume II-interpretation and applications*, 1983.

[218] P. R. Mishra, 'Remote Sensing Applications in Agriculture', *Grind GIS*, 2018. [Online]. Available: https://grindgis.com/remote-sensing/remote-sensing-applications-in-agriculture. [Accessed: 24-Oct-2018].

[219] GEOSYS, 'Understanding and Evaluating Satellite Remote Sensing Technology in Agriculture', Nova York, USA, 2016.

[220] A. Buffalano and P. Kochanowski, 'Benefits to world agriculture through remote sensing', *Acta Astronaut.*, vol. 6, no. 12, pp. 1707–1715, 1979.

[221] Bureau, '7 Benefits of Remote Sensing & GIS in Agriculture', *Agriculture Post*, 2018. [Online]. Available: https://agriculture.newsfoundry.in/7-benefits-of-remote-sensing-gis-in-agriculture/. [Accessed: 24-Oct-2018].

[222] Remote Sensing Agriculture, 'Crop Insurance/PMBFY = Remote Sensing Play Major Role in Near Future', *Remote Sensing Agriculture*, 2017. [Online]. Available: https://remote-sensing-agriculture.blogspot.com/?view=magazine. [Accessed: 24-Oct-2018].

[223] National Remote Sensing Center, 'Agriculture', *Indian Space Research Organization*, 2017. [Online]. Available: https://nrsc.gov.in/Agriculture. [Accessed: 24-Oct-2018].

[224] B. D. Wardlow, S. L. Egbert, and J. H. Kastens, 'Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains', *Remote Sens. Environ.*, vol. 108, no. 3, pp. 290–310, 2007.

[225] J. H. Kastens, T. L. Kastens, D. L. A. Kastens, K. P. Price, E. A. Martinko, and R.-Y. Lee, 'Image masking for crop yield forecasting using AVHRR NDVI time series imagery', *Remote Sens. Environ.*, vol. 99, no. 3, pp. 341–356, 2005.

[226] Z. Balint, 'Drought monitoring with the combined drought index', in *FAO-SWALIM*, 2011, pp. 3–25.

[227] G. L. Galford, J. F. Mustard, J. Melillo, A. Gendrin, C. C. Cerri, and C. E. P. Cerri, 'Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil', *Remote Sens. Environ.*, vol. 112, no. 2, pp. 576–587, 2008.

[228] T. Sakamoto, M. Yokozawa, H. Toritani, M. Shibayama, N. Ishitsuka, and H. Ohno, 'A crop phenology detection method using time-series MODIS data', *Remote Sens. Environ.*, vol. 96, no. 3, pp. 366–374, 2005.

[229] F. Rembold *et al.*, 'Using low resolution satellite imagery for yield prediction and yield

anomaly detection', *Remote Sens.*, vol. 5, no. 4, pp. 1704–1733, 2013.

[230] D. J. Mulla, 'Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps', *Biosyst. Eng.*, vol. 114, no. 4, pp. 358–371, 2013.

[231] K. Van Tricht *et al.*, 'Synergistic use of radar sentinel-1 and optical sentinel-2 imagery for crop mapping: A case study for Belgium', *Remote Sens.*, vol. 10, no. 10, p. 1642, 2018.

[232] M. Lavreniuk, G. Lemoine, and N. Kussul, 'Crop classification strategies using hybrid sentinel-1, sentinel-2 and landsat-8 data series in Ukraine', in *European Space Agency Living Planet Symposium*, 2016.

[233] A. K. Maurya, D. Singh, and K. P. Singh, 'Development of fusion approach for estimation of vegetation fraction cover with drone and sentinel-2 data', in *International Geoscience and Remote Sensing Symposium*, 2018, vol. 2018-July, no. October, pp. 7448–7451.

[234] P. Kumar *et al.*, 'Estimation of winter wheat crop growth parameters using time series Sentinel-1A SAR data', *Geocarto Int.*, vol. 33, no. 9, pp. 942–956, 2018.

[235] A. K. Whitcraft, I. Becker-Reshef, and C. O. Justice, 'A framework for defining spatially explicit earth observation requirements for a global agricultural monitoring initiative (GEOGLAM)', *Remote Sens.*, vol. 7, no. 2, pp. 1461–1481, 2015.

[236] H. Skriver, M. Thoug, and A. G, 'Multitemporal C and L band polarimetric signatures of crops', *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2413–2429, 1999.

[237] H. McNairn, J. Shang, and C. Champagne, 'TerraSAR-X and RADARSAT-2 for crop classification and acreage estimation', in *International Geoscience and Remote Sensing Symposium*, 2009, pp. 898–901.

[238] D. Bargiel, 'A new method for crop classification combining time series of radarimages and crop phenology information', *Remote Sens. Environ.*, vol. 198, no. 11, pp. 369–383, 2017.

[239] H. McNairn, J. Shang, X. Jiao, and C. Champagne, 'The contribution of ALOS PALSAR multipolarization and polarimetric data to crop classification', *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 12, pp. 3981–3992, 2009.

[240] D. Bargiel, 'Capabilities of high resolution satellite radar for the detection of semi-natural habitat structures and grasslands in agricultural landscapes', *Ecol. Inform.*, vol. 13, pp. 9–16, Jan. 2013.

[241] K. Tatsumi, Y. Yamashiki, M. A. Canales Torres, and C. L. R. Taipe, 'Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data', *Comput. Electron. Agric.*, vol. 115, no. 7, pp. 171–179, 2015.

[242] A. Dara *et al.*, 'Mapping the timing of cropland abandonment and recultivation in northern Kazakhstan using annual Landsat time series', *Remote Sens. Environ.*, vol. 213, no. 10,

pp. 49–60, 2018.

[243] H. McNairn, C. Champagne, J. Shang, D. Holmstrom, and G. Reichert, 'Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories', *ISPRS J. Photogramm. Remote Sens.*, vol. 64, no. 5, pp. 434–449, 2009.

[244] X. Blaes, L. Vanhalle, and P. Defourny, 'Efficiency of crop identification based on optical and SAR image time series', *Remote Sens. Environ.*, vol. 96, no. 3–4, pp. 352–365, 2005.

[245] G. Chust, D. Ducrot, and J. L. Pretus, 'Land cover discrimination potential of radar multitemporal series and optical multispectral images in a Mediterranean cultural landscape', *Int. J. Remote Sens.*, vol. 25, no. 17, pp. 3513–3528, 2004.

[246] J. Inglada, A. Vincent, M. Arias, and C. Marais-Sicre, 'Improved early crop type identification by joint use of high temporal resolution sar and optical image time series', *Remote Sens.*, vol. 8, no. 5, p. 362, 2016.

[247] G. Forkuor, C. Conrad, M. Thiel, T. Ullmann, and E. Zoungrana, 'Integration of optical and synthetic aperture radar imagery for improving crop mapping in northwestern Benin, West Africa', *Remote Sens.*, vol. 6, no. 7, pp. 6472–6499, 2014.

[248] S. Skakun, N. Kussul, A. Y. Shelestov, M. Lavreniuk, and O. Kussul, 'Efficiency assessment of multitemporal c-band radarsat-2 intensity and landsat-8 surface reflectance satellite imagery for crop classification in ukraine', *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 9, no. 8, pp. 3712–3719, 2016.

[249] J. L. Elman, 'Finding Structure in Time', *Cogn. Sci.*, vol. 14, no. 2, pp. 179–211, 1990.

[250] A. Afshine and A. Shervine, 'CS 230 - Recurrent Neural Networks Cheatsheet', *stanford.edu*, 2019. [Online]. Available: https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks. [Accessed: 24-Dec-2020].

[251] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, 'Gradient flow in recurrent nets: the difficulty of learning longterm dependencies', in *A Field Guide to Dynamical Recurrent Networks*, IEEE Press, 2001, pp. 237–243.

[252] S. Hochreiter and J. Schmidhuber, 'Long Short-Term Memory', *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[253] K. Cho *et al.*, 'Learning phrase representations using RNN encoder-decoder for statistical machine translation', in *Conference on Empirical Methods in Natural Language Processing*, 2014, pp. 1724–1734.

[254] M. Main-Knorn, B. Pflug, J. Louis, V. Debaecker, U. Müller-Wilm, and F. Gascon, 'Sen2Cor for Sentinel-2', in *Image and Signal Processing for Remote Sensing XXIII*, 2017, vol. 1042704, pp. 37–48.

[255] N. Brodu, 'Super-resolving multiresolution images with band-independent geometry of

multispectral pixels', *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4610–4617, 2017.

190