SYNERGISTIC USE OF MICROWAVE AND OPTICAL SATELLITE DATA FOR MONSOON CROPLAND MAPPING IN INDIA

by

Mohammad Abdul Qadir Khan

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Data Science

Spring 2020

SYNERGISTIC USE OF MICROWAVE AND

OPTICAL SATELLITE DATA FOR MONSOON

CROPLAND MAPPING IN INDIA

by

Mohammad Abdul Qadir Khan

Approved: ______________________________

Pinki Mondal, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: ______________________________

Delphis F. Levia, Ph.D.
Chair of the Department of Geography and Spatial Sciences

Approved: ______________________________

John A. Pelesko, Ph.D.
Dean of the College of Arts and Sciences

Approved: ______________________________

Douglas J. Doren, Ph.D.
Interim Vice Provost for Graduate and Professional Education and Dean of the Graduate College

ACKNOWLEDGMENTS

First, I would like to thank my committee members Dr Pinki Mondal, Dr Delphis Levia and Dr Rodrigo Vargas for their constructive feedback and intellectual support during this research work. I would like to personally offer my deepest appreciation to my advisor Pinki for her guidance for the thesis and journal manuscript, for her constant encouragement and support throughout my master's program. I would like to thank Department of Geography and Spatial Sciences for providing me funding to publish the journal paper and providing the logistics for completing my thesis on time. I would like to thank Richard Braun, Graduate Director, Masters of Science in Data Science (MSDS) program for allowing me to be part of MSDS program and providing me constant guidance for the degree. I am grateful to the officials of District Agriculture Department of Uttar Pradesh, Madhya Pradesh, Andhra Pradesh, the scientists of Acharya N G Ranga Agricultural University, Guntur and Touseef Ahmed (Acharya Nagarjuna University College of Engineering and Technology) for their assistance and co-operation during the field surveys I would also like to thank my parents and family members for their ongoing love and support throughout my academic years I appreciate constant support of Meghavi and Matt, Sindhu and Ramiz during this research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Monsoon crops play a critical role in Indian agriculture, hence, monitoring these crops is vital for supporting economic growth and food security for the country. However, monitoring these crops is challenging due to limited availability of optical satellite data due to cloud cover during crop growth stages, landscape heterogeneity, and small field sizes. In this work, our objective is to develop a robust methodology for high-resolution (10 m) monsoon cropland mapping appropriate for different agro-ecological regions (AER) in India. I adapted a synergistic approach of combining Sentinel-1 Synthetic Aperture Radar (SAR) (also called as radar) data with Normalized Difference Vegetation Index (NDVI) derived from Sentinel-2 optical data using Machine Learning algorithms of Random Forest (RF) and Support Vector Machine (SVM) within the Google Earth Engine platform. I developed a new technique, Radar Optical cross Masking (ROM), for separating cropland from non-cropland by masking out forest, plantation, and other non-dynamic features. The methodology was tested for five deferent AERs in India, representing a wide diversity in agriculture, soil, and climatic variations. Our findings indicate that the overall accuracy obtained by using the radar-only approach is 90% and 80 % whereas that of the combined approach is 93% and 90% using RF and SVM respectively It is also observed that overall RF outperformed SVM, however SVM showed improved performance when optical datasets are combined with radar data Our proposed methodology is particularly effective in regions with cropland mixed with tree plantation/mixed forest, typical of smallholder dominated tropical countries. The proposed agriculture mask, ROM, has

high potential to support the global agriculture monitoring missions of Geo Global Agriculture Monitoring (GEOGLAM) and Sentinel-2 for Agriculture (S2Agri) project for constructing a dynamic monsoon cropland mask

Chapter 1

INTRODUCTION

1.1 Background

The world population is expected to reach more than nine billion by 2050 and with increasing population, the demand for food will grow throughout the world (FAO 2017, Van der Mensbrugghe et al.2009) (Figure 1.1). The constant growing population will pose major challenges in future actions for food suppliers, policy makers and researchers. The food production are also affected by climate variability, extreme weather events characterized by less frequent and more intense rain/drought, pest infestation and many others (Rosenzweig et al. 2001). These food security concerns are more relevant for developing countries where population is projected to grow faster than their current ability of food production. The food demand is expected to increase by 70% of the current scenario until 2050 and the major increase of food production (almost double) is expected to occur in developing countries. This increasing demand of food production can only be achieved by taking necessary steps by increasing agriculture production both in developed and developing world (FAO 2017).

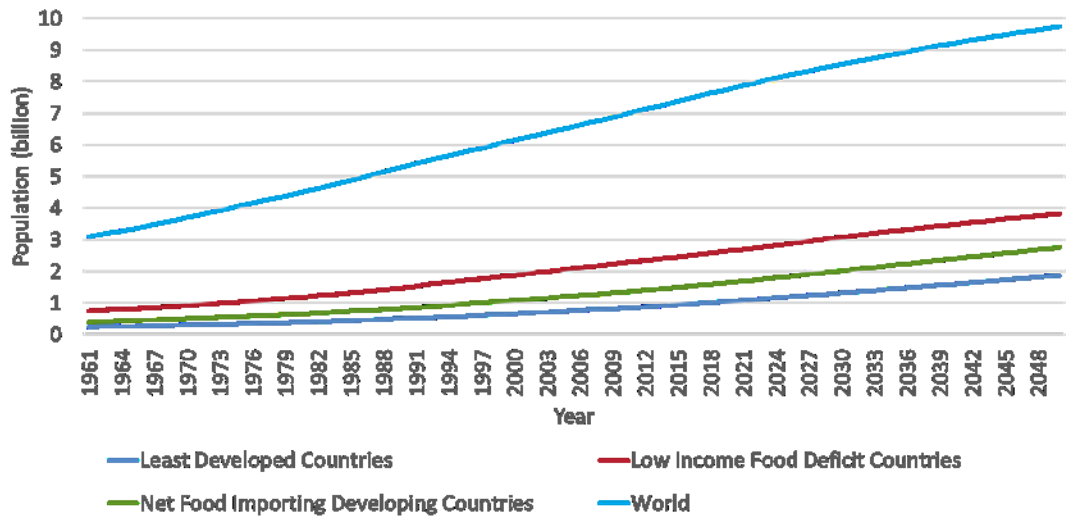

Figure 1.1: Population growth rate of the World from 1900 to 2050 scenarios (Source:FAOSTAT 2017, Van der Mensbrugghe et al. 2009)

Most of the farmers in the developing world are smallholders that own or cultivate less than 2 hectares of lands (Lowder et al. 2016, FAO 2015-2). The smallholder farmers constitute 475 million population with 28-31% of food production and constitute 24% of the gross agriculture area of the world (Ricciardi et al. 2018) (Figure 1.2). Almost 80% of the croplands in developing countries of Asia and Africa are managed by smallholder farmers (Lowder et al 2016, FAO 2015). These farmers are mainly concentrated in rural and diverse landscapes in their countries and play a vital role in livelihood creation amongst the rural population and for maintaining household food security. Their cropping intensity or yields are higher is somewhat higher than the medium and large size farms (Ricciardi et al. 2018). There is a need to support the increasing food demand by maintaining constant supply through crop production and protecting the crops grown by smallholder farmers from extreme weather events with more focus on the countries of Asia and Africa.

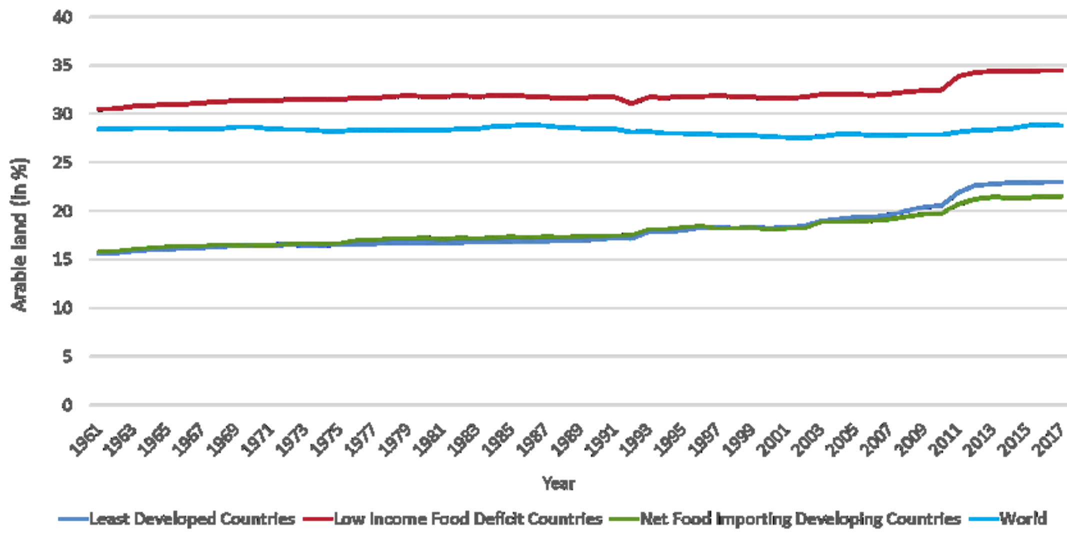

Figure 1.2: Arable land and land under permanent crops: past as percentage of total land cover (Source:FAOSTAT)

`Among the regions dominated by smallholder agriculture, India is one of the most important countries because it supports the largest agriculture population in the world and has potential to increase crop productivity (World Bank 2012). The smallholder farmers (farm size/cropland of < 2 ha or less) comprise of more than 80% of the country's farmers, but they own only approximately 25% of the total cultivated land and produce 40% of the country's food grains production (FAO 2015-2, Grain). It is estimated that as the population of the country increases the number of small farm holdings will increase throughout the country. These smallholders mainly practice rainfed agriculture and depends heavily on monsoon/wet season rainfall, thus are more vulnerable to erratic monsoon rainfall patterns.

In order to increase the crop productivity of the smallholder farmers, the first step is to understand the spatial extent of current crop cultivation area and to characterize their spatial/temporal variability across the region (Jain et al. 2013). This process will assist in generating reliable agriculture statistics and will help planning for better strategy to alleviate poverty in the region. However, these agriculture lands are changing rapidly over time and space due to anthropogenic and natural causes, which makes it very difficult to produce precise cropland maps and its geographical variations across large area (Timmermans 2017). In addition, the ground survey of these agricultural lands to generate crop estimates is not cost- or time-effective. Often these surveys do not meet the requirement of the decision-making management (Handique et al 2017).There is always a great uncertainty in generating quality cropland products, which in turn affects the local and global food security assessments. This challenge requires the need to develop new and effective methods to map and monitor the distribution of agricultural lands and crop types (crop mapping).

Remote Sensing technology can be used to extract information on the cultivated crops and cropland area in a rapid and timely manner (Debats et al. 2016). With the increasing number of earth observation satellites, the potential of remote sensing data for agriculture monitoring is increasing. New and automated remote sensing satellite-based methods are being evolved to provide agriculture related statistics (Delrue et al. 2013). Remote Sensing data has shown its importance in extracting different crop characteristics, including cropped area, crop yield, and crop damage assessment due to floods or drought, and has reduced the associated cost in conducting large field surveys (FAO 2014). These satellite-derived data products are particularly important as they can link cropping activities to the environmental factors such as soil, topography, and weather variability (Mondal et al. 2014, 2015, Qadir and Mondal 2020). With the increase in the spatial, spectral and temporal coverage of the satellites, the crop mapping products are also improved over time (Jin et al. 2019). However, in tropical countries like India, crop monitoring during the monsoon season is still a challenging task as cloud free optical scenes are difficulty to get to generate reliable crop statistics (Qadir and Mondal 2020, Singha et al. 2019). Moreover, the field sizes are very small to be efficiently mapped by existing methods (Jain et al. 2013).

1.2 Cropland mapping using remote sensing

Historically, earth observation using remote sensing been used for crop mapping and area assessment since the launch of Landsat satellite in 1972 (MacDonald et al.1975, Wulder et al. 2019**)**.With the advancement of technology and sensor development, satellite remote sensing has become a stronger player in many countries for operational crop monitoring work. Initially, US organizations including United States Department of Agriculture (USDA), National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), United States Department of Commerce (USDC) coordinated to carry out a joint "Large Area Crop Inventory Experiment" program (MacDonald et al. 1975). New programs were started not only to map major food crops (wheat, rice, soy) but also to assess the crop production estimation of major crop producing countries of the world.

Initially, the satellite data from Advanced Very High Resolution Radiometer (AVHRR) and Landsat was the main source to generate agriculture statistics by using vegetation indices in the USA (Khamala 2017). New models for estimating crop condition and biomass were then developed by using remote sensing and meteorological data as collaborative work between the European Unions and the USA (Fritz et al. 2019).Other countries followed to start their own program for crop mapping and production estimation including India. However, most of the prior case studies at the national or global scale were implemented using coarser satellite data such as Moderate Resolution Imaging Spectroradiometer (MODIS) or AVHRR (Jiang et al. 2003).

Methodologies involving such coarser data, when applied to small-scale agriculture (farm sizes/cropland less than 2 hectares), common among transitioning economies, result in mixed pixel issues where one aggregated grid-cell value is assigned to many fields with varying cropping practices (Qadir and Mondal 2020). Using coarse resolution data is not efficient in making accurate classification of all land uses including croplands with the required standard of accuracy (Handique et al 2016). With the availability of large amount of satellite data, improvement in resolution of the available satellite imagery (radiometric, temporal and spatial), and advancement in machine learning techniques, regular cropland mapping and monitoring are becoming more common in both science and policy sectors (Mtiba and Irie 2016).

For efficient cropland monitoring, the timing or schedule of crop area estimation depends on how early and efficiently the planted crop can be detected by satellite sensors. As crop growth is dynamic, it was demonstrated that multi-temporal data could significantly improve the crop mapping accuracy (Sun et al. 2019). It also depends on the field sizes, the spatial variability of the growing crops and the timing of the crops grown. For example, in India where the rainfed crops are grown in monsoon (wet) season, obtaining cloud free optical data is challenging, resulting in severe limitation in monsoon crop mapping. With the availability of synthetic aperture radar (SAR) sensors such as Sentinel-1 and ALOS-PALSAR, monsoon crops can now be monitored throughout the crop growth season (Van Tricht et al. 2018). However, from the operational monitoring context, these new mapping techniques must be replicable over different agro-ecological regions covering large geographic extent and should be robust enough to be applicable for different years for the same crop-growing season.

1.3 Cropland mapping using remote sensing: Indian scenario

India is a primarily agrarian economy with 17% of the national Gross Domestic Product (GDP) contributed by agriculture and approximately 50% of the population supported by agricultural activities (Madhusudhan 2015). The remote sensing activities in country began with the study of root-wilt disease in coconut plantation in Kerala (Dadhwal et al 2002). National level crop area estimation programs such as Crop Acreage and Production Estimation (CAPE) and Forecasting Agricultural Output using Space, Agro-meteorology and Land based Observations

(FASAL) also became operational (Parihar et al. 2016).The aim of FASAL program was to generate crop forecast using remote sensing data at an early stage of crop development and to regularly update the existing forecast regularly to improve the ground based crop forecasting (Moorthi et al. 2014). The FASAL procedure is now operationally used by Mahalanobis National Crop Forecast Centre (MNCFC) for national level crop area estimates. In FASAL, cloud free multi-spectral optical remote sensing data from medium resolution satellites such as Wide Field Sensor (WiFS) and Advanced Wide Field Sensor (AWiFS) are used for crop monitoring over large regions for both winter/dry (*Rabi*) and monsoon/wet (*Kharif*) crops. These mapping activities include: (1) extracting spectral signatures and indices including Normalized Difference Vegetation Index (NDVI) using optical satellite data, and (2) assessing temporal evolution of various crops and other land cover features using sample segments collected from the ground. Based on these sample segments, a classifier model would generally be used at the district and state level. These state level statistics were combined to provide national level statistics. Also, India is one of the few countries who have initially used SAR data for operational monsoon (*Kharif*) crop mapping for selected crops when optical data availability is limited. India launched its own C- band SAR satellite Radar Imaging Satellite (RISAT-1) in April 2012 for boosting its FASAL and monsoon crop mapping program (Chakraborty et al. 2014). Before RISAT-1, temporal C-band Radarsat-1 & 2 were explored for water intensive rainfed monsoon crops, such as rice and jute. Before RISAT-1, temporal C-band

Radarsat-1 & 2 were explored for water intensive rainfed monsoon crops, such as rice and jute.

However, most of the studies conducted in India on small farms using moderate–high resolution satellite data such as Landsat and Sentinel-2 mainly focus on winter crops when sufficient cloud free optical data are available during the crop-growing season (Whitcraft et al. 2015) However, optical satellite data are insufficient for operational monsoon cropland mapping as the wet (monsoon) season coincides with the crop growing duration, thus, providing an insufficient number of images for mapping monsoon cropland over a large scale. Even when optical satellite data are available during the peak growth stages of the crops such as rice, the spectral signatures of the crops are often mixed with that of plantation, grassland, or forested regions (Mercier et al.2019), thus making it challenging to segregate croplands with monsoon crops from other vegetation covers.

1.4 Machine learning based classification for cropland mapping

Classification of satellite imagery is an important component for land cover studies comprised of change detection analysis, management of resources and crop damage preparedness. Up-to-date information of land cover is required for policy implementation for future conservation projects (Roy and Inamdar 2019). Cropland classification can be performed by either using pixel-based information such as those used in supervised/unsupervised classification algorithms or by clustering similar objects together in an object-based classification. Pixel-based classification algorithm is easier to implement as it requires less computational resources compared to object-based classifications (Quynh Trang et al. 2016). In supervised classification, the classifier is calibrated based on the initial training values and the selected algorithm, whereas in unsupervised classification, an algorithm is selected that will find pre-specified number of clusters within the input image based on statistical analysis (Macedo-Cruz et al. 2011). Unsupervised classification doesn't require prior ground information for generating land cover classes. More recently, with the advancement of computation power and development of new algorithms, satellite image processing using machine learning has evolved as the demand has increased to have a smart tools and software to detect the hidden pattern behind the satellite images (Debats et al. 2016).These algorithms range from Artificial Neural Networks (ANN), k-Nearest Neighbors (kNN), Decision Trees (DT), Classification and Regression Trees (CART), Support Vector Machines (SVM), and Random Forest (RF) (Qian et al. 2014, Salah 2017, Okwuashi et al.2012). A comparison of commonly used Machine learning techniques is given in the table below 1.2 below

Machine learning which uses statistical learning to train the model having a set of features or attributes differ from Deep Learning (DL) which is a subset of Machine learning such that it extracts features or attributes from raw data (Table 1.1). Deep learning is characterized by neutral networks generally involving more than two hidden layers, hence the term deep learning which require powerful computational resources. Some of the widely used deep learning models are Convolution neutral Network (CNN) and Recurrent Neural Network (RNN) (Ma et al. 2019).

Table 1.1: Comparison of Machine Learning (ML) vs Deep Learning (DL)

Factors	**Machine Learning**	**Deep Learning**
Data Structure	Always require structured data	Uses layers of neural networks
Data size	Can train on lesser data	Require large data
Execution time	Less time compared to DL	More time compared to ML based on number of parameters used
Accuracy	Less accurate	Provides high accuracy
Hardware	Can be trained on Central Processing Unit (CPU)	Requires high quality Graphics Processing Unit (GPU)

Though the machine learning classifiers are widely gaining acceptance due to their high accuracy, their potential for mapping small-scale farms/croplands are still not explored properly. Very few studies have been conducted, but none of them explores its potential for monsoon cropland mapping. There has been limited amount of research to compare and evaluate the performance of machine learning algorithms, such as RF and SVM with combination of radar and optical data for monsoon cropland mapping. Previous studies have shown that these algorithms are very sensitive to the training dataset and the algorithms parameters used (Thanh Noi and

Kappas 2018). For example, decision trees are too sensitive to small changes in the training dataset and occasionally it is unstable and tend to overfit the model (Topaloglu et al 2016). Studies have also shown that SVM and RF are insensitive to noise or overtraining, which shows their ability in dealing with unbalanced data (Breiman 2001).

With the availability of cloud computing platform such as Google Earth Engine (GEE), and accessibility to several in-built machine learning algorithms, it has now become easier to analyze large number of satellite imagery (Gorelick et al. 2017). GEE supports more than 15 classification techniques including machine learning algorithms such as SVM, RF and CART. Two of the algorithms (SVM and RF) supported by GEE and used for this study are explained in detail below. GEE supports more than 15 classification techniques including machine learning algorithms such as SVM, RF and CART. Two of the algorithms (SVM and RF) supported by GEE and used for this study are explained in detail below.

1.4.1 Support Vector Machine (SVM)

SVM is a supervised, non-parametric machine-learning algorithm that can be used for both regression and classification. The popularity of SVM has increased recently in the classification of satellite images. In SVM, a hyper-plane is found and fit on the data set in N-dimensional space (where N is number of features) that separates the training classes in feature space (Figure 1.4). In SVM, the training points play an important role in generating highly accurate classification. The easiest way to train the

SVM is by using linearly separable classes. According to (Mountrakis et.al.2011,Thanh Noi and Kappas 2018) if the training data with k number of samples is represented as {Xi, Yi}, i =1, 2,..., k where X € Rn is an n-dimensional space and y € {−1, +1} is a class label, then these classes are considered linearly separable if there exists a vector W perpendicular to the linear hyper-plane (which determines the direction of the discriminating plane) and a scalar b showing the offset of the discriminating hyper-plane from the origin. For the 2 classes, i.e. class 1 represented as −1 and class 2 represented as +1, 2 hyper-planes can be used to discriminate the data points in the respective classes. These are expressed as;

$WXi + b \geq +1$ for all $y = +1$, i.e. a member of class 1

$WXi+b \leq -1$ for all $y =-1$, i.e. a member of class 2

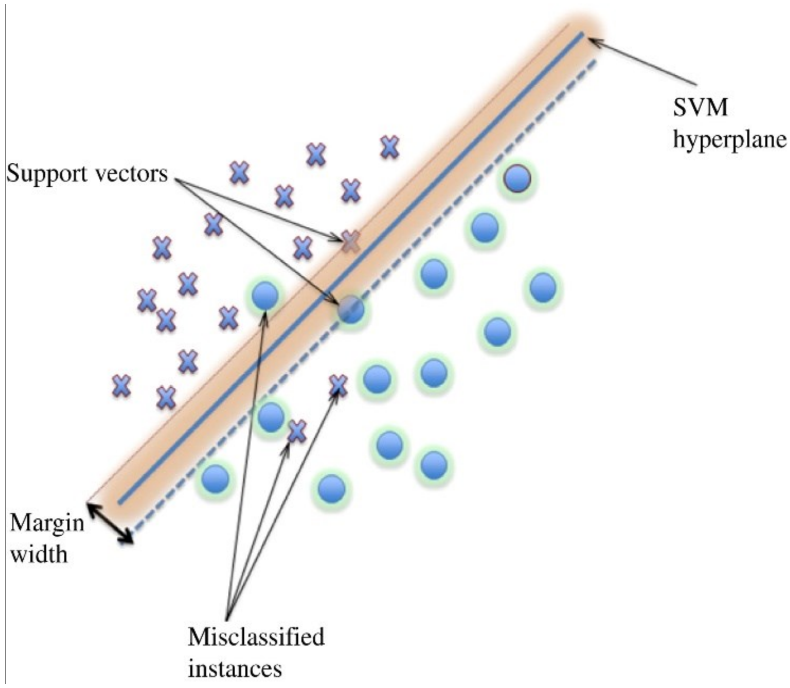

Figure 1.3: The basic principle for Linear Support Vector Machine Classifier (Source: Mountrakis et.al.2011)

In some cases, the classes might not be linearly separable which is often the case in land cover studies. In the case of non-linearly separable classes, to find the optimal boundary, the training points are to be projected in a higher dimensional space where the data points become linearly separable (Huang et al. 2002). For these kind of classification tasks, kernel representations offer a solution in locating complex boundaries between the classes. There functions (kernels) take into account the low dimensional input space and convert it into high dimensional space. The SVM classifier provides four types of kernels: linear, polynomial, radial basis function

(RBF), and sigmoid. SVM is regularly used for agriculture mapping at a larger scale, such as for the Sen2Agri project, where the automated crop type mapping was performed over 12 sites across the world using time series optical images coupled with RBF kernel based SVM classification (Inglada et al. 2015).

1.4.2 Random Forest (RF)

RF is another classification algorithm, which follows the decision tree approach. In RF, randomly selected results from multiple decision trees are combined together to obtain highly accurate and stable classification results (Breiman 2001) (Figure 1.5**)**. Similar to SVM, RF can also be used for both classification and regression problems. According to (Thanh Noi and Kappas 2018) in order to implement the RF, two parameters need to be set up: the number of trees and the number of variables per split. GEE default values for the above two parameters (i.e. number of trees equal to 1 and number of variables per split is equal to square root of the number of variables) optimum results can be achieved (Pal 2005, Liaw et al.2002), whereas changing the two parameters may or may not improve the performance of the classification accuracy depending on the input image and the training points used for the algorithm.

Some studies have shown that by increasing the number of trees to 200, RF could achieve accurate results (Feng et al. 2015, Inglada et al. 2015, The objectives of this study is to evaluate the performance of the supervised classifiers, RF and SVM, when applied to a Sentinel-1 radar and combination of Sentinel-1 radar and Sentinel-2 optical satellite image and to assess the accuracy of the classification results.

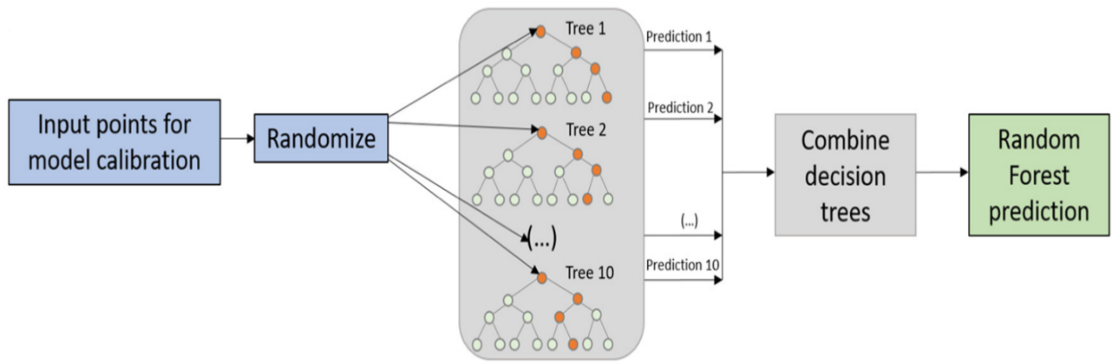

Figure 1.4: Methodological framework for Random Forest classifier depicting the internal structure of a Random Forest (Source: Mondal et al. 2020)

1.4.3 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is a technique, which imitate the power of human brain to perform tasks such as pattern recognition, image classification etc (Maxwell et al. 2018). It differs from Random Forest or other conventional satellite image classifiers, as it is not based on decision rules (Lange and Sippel, 2020). Previous studies have shown considerable advantages of ANNs over the conventional methods for satellite image classification due to its ability to learn complex patterns taking into account any non-linear complex relationship between the dependent and independent variables. ANN also take into account any prior knowledge while gained while training the classifiers. In ANN, the neuron forms the basis of an ANN where each neuron is connected and require certain weight for activation or inhibition of the neuron The efficiency of the neural networks is based on the optimization of the learning algorithm, the parameters chosen for training and the quality of the data used. ANNs is now been applied to image classification, feature extraction, data fusion and other similar tasks. Nowadays, Neural Network has gain popularity for satellite image classification due to improvement in the algorithms and the processing systems

Table 1.2:Comparison of the three algorithms: Support Vector Machine (SVM), Random Forest (RF) and Neural Networks (NNs)

Algorithms	**Advantages**	**Disadvantages**	**User defined parameters**
Support Vector Machines	• Order of the instances doesn't matter. • Overfitting rarely a issue	• Processing time increases exponentially as the classes increases • Choosing appropriate hyper parameter is tough	• Cost parameter • Kernel type : linear, polynomial, radial
Random Forest	• Order of the instances doesn't matter • Tolerate complex data • Easy to optimize	• Good for classification task but not for regression • Overfitting problem	• Number of trees • Number of variables per split
Neural Network	• Tolerate noisy data • Able to represent Boolean functions	• May lead to overfitting • Structure of the algorithm is complex and not easy to understand	• Activation function • Number of hidden layers • Learning rate • Pruning parameter

1.4.4 Crop monitoring using multi-temporal radar data

The radar data operates on low frequency (1-10 GHz) enabling penetration of the cloud cover and solar independence (Woodhouse 2006). The radar observations in this frequency range are also sensitive to soil moisture content and roughness, vegetation size, shape, orientation and biomass. Hence, the low frequency, along with the sensitivity to soil moisture and plant structural properties make the radar sensor more suitable for crop monitoring. Previous studies have shown that utilizing single SAR scene at a given frequency or polarization is often inadequate to achieve the desired classification accuracy (Skriver et al. 2011). Studies have also shown how temporal backscattering can efficiently differentiate crops based on their canopy and other physical attributes. (Shang et al. 2009).

Multi-temporal SAR images improve the crop classification accuracy and capture the variation in growth process (Larrañaga and Álvarez-Mozos 2016). Chakraborty et al. 1997 has documented the use of multi-temporal SAR data for classifying agricultural lands and monitoring crop growth. Skakun et al. 2016 has shown how multi-temporal SAR images can effectively produce the equivalent classification accuracy as optical images during the cloudy seasons. Studies have also shown that multi-temporal SAR images (>10 scenes) can increase the classification accuracy obtained from optical images (2 or 3 scenes) by 5% (Kussul et al.2018). Skriver et al.2011 has shown that multi-temporal, multi-polarization SAR images perform better compared to single date multi-polarization or multi-date single polarization radar images. Hence, the temporal information from SAR, combined with multi-polarization, provide better information of crop conditions. Nowadays, open source Sentinel-1 radar data with high spatial resolution of 10m, high temporal revisit frequency (10 days) and C- band in dual polarimetry is widely used for crop monitoring. Studies have shown that C-band interact more efficiently with crops due to its lower penetration capability compared to L-band and less canopy scattering as compared to X-band (Inoue et al. 2002). Wide range of studies are performed for cropland monitoring and yield estimation by utilizing backscattering values of single or dual polarimetric Sentinel-1 radar data. Currently, water intensive crops, such as rice, are widely monitored by radar data (Haldar et al. 2014, Mansaray et al. 2017, Rakwatin et al. 2014, Singha et al 2016, 2019). However, few studies have utilized radar data for dryland crops grown during wet season. Previous studies utilizing radar have been confined to examining croplands dominated by specific water intensive monsoon crops such as rice or jute which are easier to detect due to their distinct backscattering signatures compared to dryland monsoon crops (Wang et al. 2015). Hence, these radar-based methods need to be evaluated or revised in the context of diverse cropping practices, especially for rainfed monsoon crops grown in dryland regions. The hindrance in using optical satellite data for intra-seasonal monsoon cropland monitoring over large region requires the remote sensing community to develop new methods, especially for countries with heterogeneous landscapes, such as India.

1.4.5 Synergy of radar and optical data for crop monitoring

Multi-sensor combination of satellite data capturing different parts of electromagnetic radiation provides us the possible strategies to reduce the effect of cloud covers. Many of the operating microwave satellites such as Sentinel-1, ALOS PALSAR operates in C-band (3.75-7.5 cm), L-band (15-30 cm) wavelengths respectively which are not affected by clouds and heavy rainfall. Hence, they are ideal for crop monitoring during the monsoon season in India. However, these satellites have their limitation of interpretation and speckle effects. In addition, interaction of radar with dryland regions is limited. By combining radar and optical data for the crop growing season, advantages of both sensors can be maximized, while limiting the drawbacks in using data from either of these sensors as a single input. Moreover, by using the time series imagery, the temporal variations of the crop growing cycle can easily be captured and data gaps resulting from non-availability of satellite data are avoided. McNairn et al. 2009 integrated the Landsat TM with Radarsat and ENVISAT in Canada and concluded that if even one optical images are combined with temporal radar images than acceptable accuracy of (>85%) can be achieved for operational purposes. Previous study conducted by (Inglada et al. 2016) has shown that there is significant improvement in classification accuracy early in crop growing season by fusing time series of Landsat and Sentinel-1 radar data.

Similarly Shang et al. 2008 achieved acceptable accuracy (>85%) in early growing season by using Landsat data complemented by multi-polarization (VV/VH) ENVISAT Advanced Synthetic Aperture Radar (ASAR) images. Studies have shown that there is always an increase in classification accuracy ranging from 5% to 25% when the two data sources are combined (e.g. Kussul et al. 2018; Skakun et al 2016).

Recently, during monsoon crop growing season, studies have emerged to combine radar and optical data for predicting yields of several monsoon crops including rice, soybean, and cotton (Ranjan et al. 2019, Kumari et al. 2019). These studies either focused on water intensive monsoon crops or over small regions where obtaining a few optical image snapshots was possible during the monsoon season. However, no studies thus far integrated radar and optical data for quantifying monsoon cropland over a large area in different agro-ecological regions with diverse agriculture systems.

1.5 Research Objectives

This study intends to fill the gap in monsoon cropland mapping by combining radar and optical data and has the following objectives:

(1) Evaluating Sentinel-1 (S1) radar and a combination of radar and Sentinel-2 (S2) optical data in terms of providing greater accuracy for monsoon cropland mapping using machine learning algorithms.

(2) Developing a high resolution, all weather applicable non-crop mask for segregating monsoon cropland from forested and agro-forested (plantation) lands with similar signatures.

The above research objectives can be achieved by addressing the following research questions:

- How well do the in-built machine learning methods of GEE such as RF and SVM perform on multi-temporal satellite images for monsoon cropland mapping?
- How do radar data alone and combination of radar with optical data perform across different AERs by using RF and SVM classification?
- How well do these two classifiers perform with respect to each other and what is the pixel level agreement (PLA) between RF and SVM classified images?

Chapter 2

STUDY AREA & DATASETS USED

2.1 Study Area

The study area comprises of ten sub-regions within five agro-ecological regions (AER) covering the Indian states of Uttar Pradesh, Madhya Pradesh, Chhattisgarh, Maharashtra, Andhra Pradesh, and Karnataka (Figure 2.1) (Gajbhiye and Mondal 2000). The study area covers approximately 604,615 sqkm and is surrounded by the alluvial Gangetic plains in the north and the Bay of Bengal in the south. It borders the Western Ghats in the west and Chota Nagpur plateau in the east. The region is mostly undulating with the elevation ranging between 0–1560 m (Figure 2.1). The region is irrigated by many rivers including the major rivers such as Narmada, Godavari and Krishna (Khullar 2008). The study region has varying elevation from the lowlands in the coastal plains to the plateau in the central to highlands on the northern side. This study area was selected as the farmers mainly practice rainfed monsoon crops and it contains great diversity of agricultural landscapes and characterized by different agro-ecological regions.

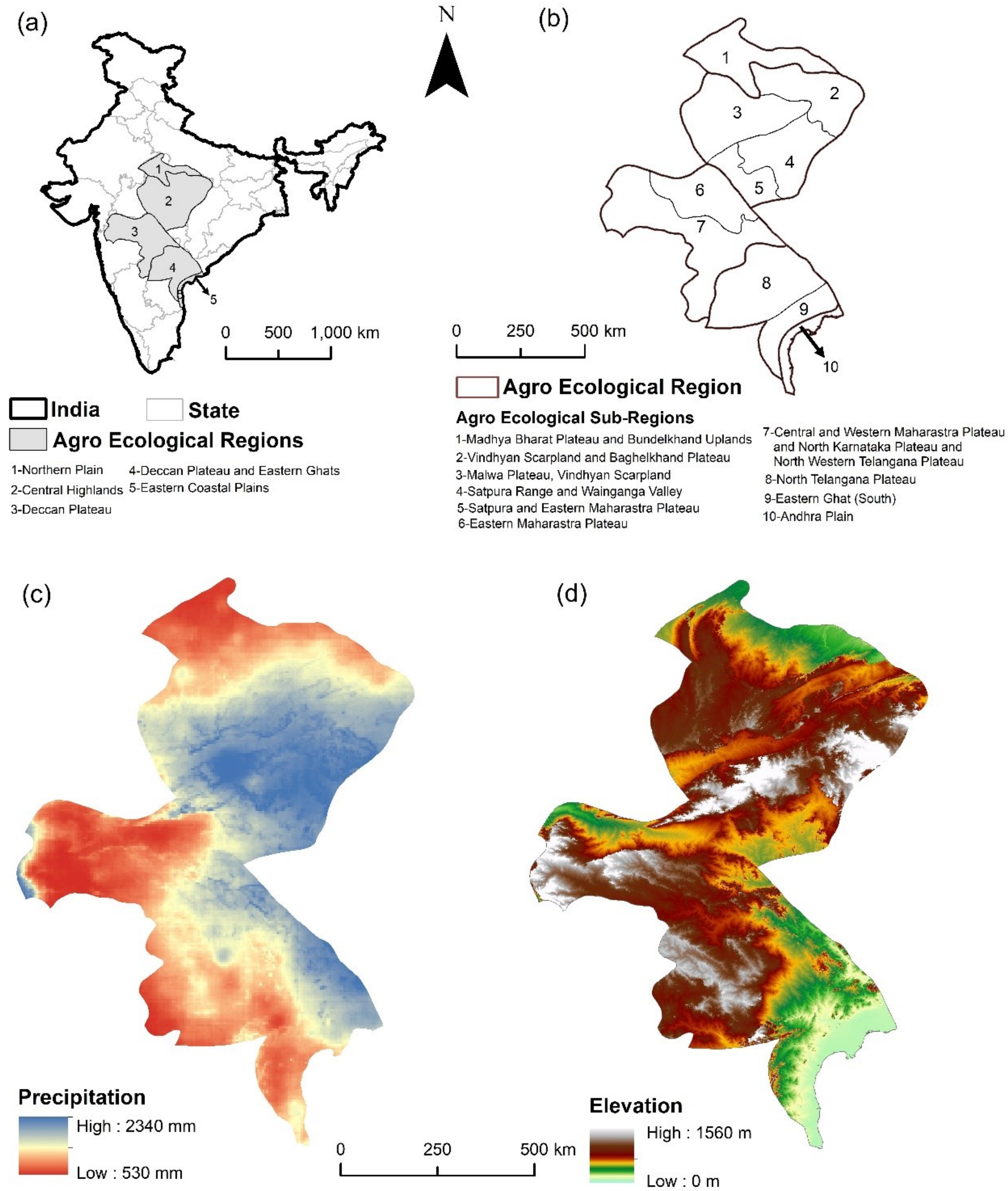

Figure 2.1: Maps of the study area showing: (a) Agro-Ecological Regions (AER) selected for this study; (b) ten AER sub-regions within five AER; (c) spatial variation in annual mean precipitation from the year 2000 to 2018, derived from the Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) data; and (d) Digital Elevation Model (DEM) obtained from the Shuttle Radar Topography Mission (SRTM) dataset.

2.1.1 Climate

The geography and the topography of India strongly influence its climate. The climate of the region is mainly characterize by wide range of weather conditions and topographical variations. The Himalayas in the north, the Thar Desert in the west and the oceans, the Indian Ocean, the Bay of Bengal and Arabian Sea has a great role in influencing its climate. The region has a tropical monsoon climate (Am) as per the Koppen-Geiger climate classification system (Beck et al. 2018). The mean monthly rainfall for the study region is shown in Figure 2.1c and mean monthly temperature (calculated from MODIS) and rainfall (calculated from Climate Hazards Group InfraRed Precipitation with Station (CHIRPS)) for the whole of India is shown in Figure 2.2, for 2000–2018.

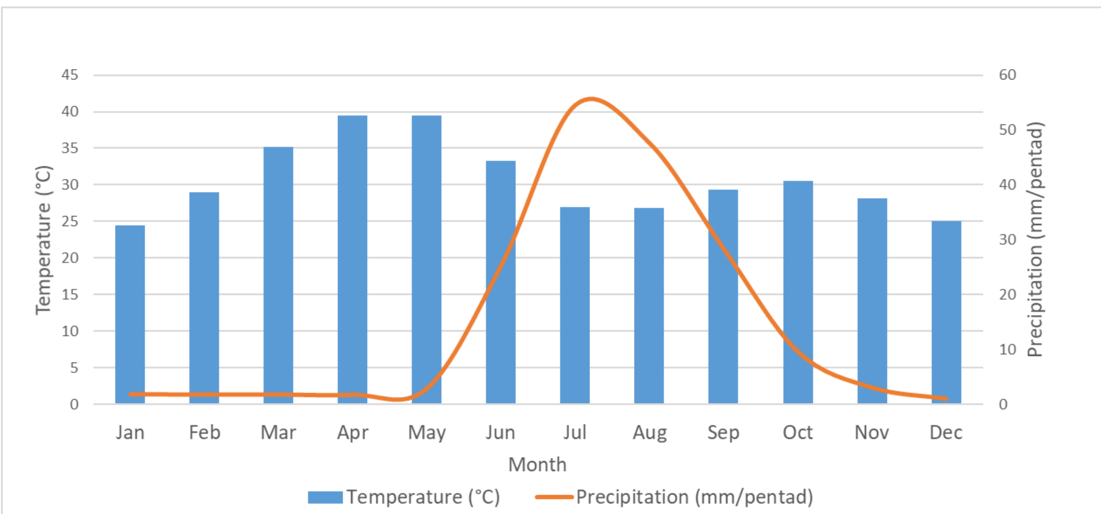

Figure 2.2: Average precipitation and rainfall for part of the study region over a year

- In India there are four climatological seasons as declared by the Indian Meteorological Department (Attri and Tyagi, 2010)
- **Winter**, occurring between January and March.
- **Summer** or **pre-monsoon** season, lasting from March to June
- **Monsoon** or **rainy** season, lasting from June to September. This season brings the summer monsoon rainfall over India
- **Post-monsoon** season, lasting from October to December. cloudless.

2.1.2 Monsoon season

Monsoon season is the main crop-growing season in India and runs for four months from June to September and dominates by massive convective thunderstorm (Parthasarathy et al. 1994, Prasanna 2014). During the monsoon season, large area of western and Central India receives more than 90% and southern and eastern India receives almost 50%-75% of their total precipitation during the period (Halpert and Bell 1996). Monsoon is derived from the word “Mausam” meaning seasonal rain bearing winds (Zhisheng et al. 2015). The monsoon season is caused by the difference in temperature over the land and the ocean resulting in difference in pressures: low pressure over land surface and high pressure over the oceans. This give rise to movement of moisture-laden trade winds from ocean to land surface and termed as monsoon (Gallup and Riker-Coleman 2001). By first week of July the wind covers the whole of India experiencing the monsoonal rainfall. By the last week of August/first week of September, the monsoon wind starts retreating from mainland India. It further weakens by the end of September/early October and leaves the entire country by the end of November.

During the months from October to December, the monsoon winds become weaker and starts retreating. This retreating monsoon carries wind that has already lost their moisture while crossing land and brings with it the cool, dry air masses to large part of India (Galvin 2008).

2.1.3 Agro-ecological regions

Agro-ecological regions (AER) is extracted from Agro-climatic regions by overlaying landforms and soils on climatic regions based on length of agricultural growing period (Gajbhiye and Mondal 2000).(Table 2.1) The AER is designed to attaining the optimum production potential of a crop and crop variety. The latest AER map available for India was prepared from using the soil resource data acquired from 1:1 million-scale soil map and climate data from 600 weather stations spread across the country. The soils in the region are influenced by the rainfall, the existing parent rock material and topography. The AER regions used in this study were prepared by laying emphasis on soil quality parameters to prepare accurate agricultural land use plans.

Table 2.1: Primary characteristics of the study region showing agro-ecological regions and sub-regions, climate and the soil types

	Agro-ecological region	**Sub-region**	**Climate type**	**Soil type**
1	Northern plain (and central highlands) including Aravallis, hot semi-arid eco-region	Madhya Bharat plateau and Bundelkhand uplands	Hot moist semi-arid	deep loamy and clayey mixed red and black soils
2	Central Highlands (Malwa and Bundelkhand), hot subhumid (dry) eco-region	Vindhyan scarpland and Baghelkhand plateau	Hot dry subhumid	deep loamy to clayey mixed red and black soils
3	Central highlands (Malwa and Bundelkhand), hot subhumid (dry) eco-region	Malwa plateau, Vindhyan scarpland and Narmada valley	Hot dry subhumid	medium and deep clayey black soils (shallow loamy black soils as inclusion)
4	Central highlands (Malwa and Bundelkhand), hot subhumid (dry) eco-region	Satpura range and Wainganga Valley	Hot moist sub-humid	shallow to deep loamy to clayey mixed red and black soils
5	Central highlands (Malwa and Bundelkhand), hot subhumid (dry) eco-region	Satpura and eastern Maharastra plateau	Hot dry sub-humid	shallow and medium laomy to clayey black soils (deep clayey black soils as inclusion
6	Deccan plateau, hot semi-arid eco-region	Eastern Maharastra plateau	Hot moist semi-arid	medium and deep, clayey black soils (shallow loamy, to clayey black soils as inclusion
7	Deccan plateau, hot semi-arid eco-region	Central and western Maharastra plateau and north Karnataka plateau and north western Telangana plateau	Hot moist semi-arid	shallow and medium loamy to clayey black soils (medium to deep clayey black soils)

8	Deccan plateau (Telangana) and eastern ghats, hot semi-arid eco-region	North Telangana plateau	Hot moist semi-arid	deep loamy and clayey mixed red and black soils
9	Deccan plateau (Telangana) and eastern ghats, hot semi-arid eco-region	Eastern ghat (south)	Hot moist semi-arid/dry subhumid	medium to deep loamy to clayey mixed red and black soils
10	Eastern coastal plain, hot subhumid to semi-arid eco-region	Andhra plain	Hot dry subhumid	deep, clayey coastal and deltaic alluvium-derived soils

2.1.4 Major crops

Most of the farmers in the study region are smallholders with limited landholdings; they grow crops during three seasons: monsoon (*kharif*) during June-November, winter (*rabi*) during December-April, and summer (*zaid*) during April–June (NFSM Report). The major monsoon crops grown in the study region are rice, soybean, black gram (locally known as *Urad*), cotton, maize, and groundnut (Land Use Report). The monsoon crop sowing date varies across the study region, starting in the month of June with the onset of monsoon, up to August/September in low-lying regions. The harvesting of the crops widely varies as well and may range from September for soybean and black gram to November for rice. The details regarding the AERs considered for this study and the major monsoon crops grown according to the latest government statistics available are listed in Table 2.2. For further analysis, we have combined AER-5 with AER-4, as AER-5 has negligible cropped area to be analyzed as separate unit. Diverse monsoon crops are grown in the AERs comprising of both water intensive monsoon crops and rainfed-dryland crops.

Table 2.2: Agro-Ecological regions and the major crops grown

	Agro-Ecological Region	Major Crops
1	Northern Plain	black gram, millet, sesame, rice
2	Central Highlands	soybean, rice, cotton
3	Deccan Plateau	cotton, soybean, sorghum
4	Deccan Plateau and Eastern Ghats, Eastern Coastal Plains	rice, cotton, chili, maize

2.2 Datasets used

2.2.1 Sentinel-1 radar data

Sentinel-1 radar data is part of Global Monitoring of Environment and Security (GMES), a combined initiative of European Commission and European Space Agency and also part of Copernicus program. Sentinel-1 consists of constellation of two satellites Sentinel-1A launched on April 3rd, 2014 and Sentinel-1B launched on April 25th, 2016. Sentinel-1 operates on C-band at dual polarization 10m spatial resolution and 12- days repeat cycle. It has four modes of imaging: strip map (SM), interferometric wide swath (IW), extra wide swath (EW) and wave (WV) mode. Sentinel-1 radar data is collected with several different resolutions, polarization combination during both ascending and descending orbits. For this study, I have used C-band, dual-polarization VV (Single co-polarization, vertical transmit/vertical receive) and VH (Dual-band cross-polarization, vertical transmit/horizontal receive) dataset for the Interferometric Wide Swath (IWS) mode in descending look angle accessed and processed on the Google Earth Engine (GEE) platform. The dataset includes the Sentinel-1 ground range and ortho-rectified product processed using the sentinel-1 toolbox and converted to backscattering coefficient in decibel (dB) scale. I filtered the data, restricted the processing and data analysis to Level-1 GRD products and sorted by dates within the crop growing monsoon season from June to November. Pre-processing included the steps for thermal noise removal, radiometric calibration and ortho-rectification of the dataset. The ortho-rectification was performed using the Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) within the GEE environment (Qadir and Mondal 2020). Using temporal VH and VV polarization, radar monthly composite images were created by considering median values. I also used these monthly median composite images to create False Color Composite (FCC) to aid in visual interpretation of the images for training and testing data collection (Figure 2.5). A total of 516 Sentinel-1 radar images were used for the entire monsoon crop-growing season of 2018.

2.2.2 Sentinel-2 optical data

The Sentinel-2 (S2) mission is a constellation of two polar-orbiting satellites similar to Sentinel-1. Sentinel -2 is wide swatch (290 km), high spatial resolution (10m) and high revisit time (10 days at the equator) data provided by ESA and supporting Copernicus Land monitoring programs. I have used the S2 level 1-C, ortho-rectified and geo-referenced top-of-atmosphere (TOA) reflectance data product

(Gati and Bertolini 2015) within the GEE platform. The collection contains Multi Spectral Instrument (MSI) bands with a scaling factor of 10,000. To maintain the quality of the data analysis and products during the monsoon season, I considered Sentinel-2 images with cloud cover of 5% or less. On these filtered images, I applied an automated cloud masking algorithm using quality assessment band (band QA60) to mask both opaque and cirrus clouds (Carrasco et al. 2019). The images acquired after the month of November were not considered as I assumed that crops grown after this month are not monsoon crops, based on existing literature (Jain et al 2013). A total of 1734 S2 images were used for the entire monsoon crop-growing season of 2018.

2.2.3 Training and Testing the Classifiers

I collected a total of 1500 reference points required for training and testing the classifiers for the five broad land use land cover (LULC) classes: monsoon crop, bare soil, water, vegetation, and urban (Figure 2.4). I defined bare soil as any land cover feature, which is devoid of vegetation, either a barren land, fallow land with no crop, or any region with exposed soil. I collected these points through a combination of field visits, high-resolution google earth imagery and visual interpretation of Sentinel-1 radar and Sentinel-2 optical satellite imagery using the method similar to those explained in Singha et al. 2019 and De Alban et al. 2018 and demonstrated in Figure 2.4. Using multiple datasets to generate the training and testing points ensures that only land cover features, which have the high probability of belonging to the actual land cover class on the ground, are selected for training and testing. During field visits, information on land covers along with their geographic coordinates Ire collected using a handheld Global Positioning System (GPS) device. Field visits were conducted at four agro-ecological sub-regions: 1) Madhya Bharat Plateau and Bundelkhand Uplands, 2) Vindhyan Scarpland and Baghelkhand Plateau, 3) Eastern Ghats (South), and 4) Andhra Plain (Figure 2.1a, Table 2.1). I collected a total of 500 points for monsoon crops, 300 points each for bare soil and vegetation, and 200 points each for water and urban, using stratified random sampling approach. The number of points collected for each land cover was decided based on the relative dominance of these land covers in the study landscape.

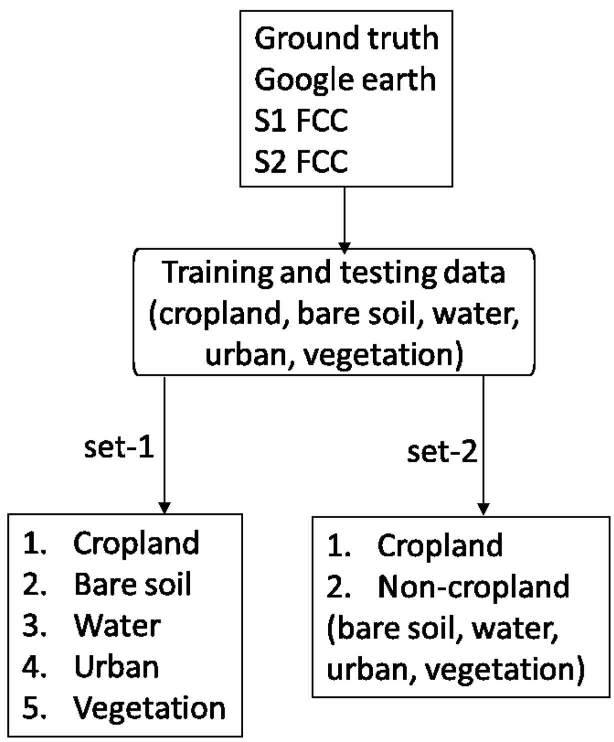

Figure 2.3: Workflow detailing the steps for collecting the training and testing points and the classes used for set-1 and set-2 using ground truth, google earth, Sentinel-1 (S1) False Color Composite and Sentinel-2 (S2) False Color Composite (FCC)

First, the field data points were imported on GEE platform and were overlaid on the Sentinel-1 (S1) radar and Sentinel-2 (S2) optical False Color Composite (FCC) images. Using this ground truth data, the extracted features were used to identify similar LULC features in other regions using visual interpretation techniques on FCC of S1 and S2. The extracted features were verified using the high-resolution google earth imagery. Extracting training points for water, forest/plantation and urban is straightforward in S1 as they are not dynamic over time and have very distinct temporal backscattering signatures compared to crops and bare soil as shown in Figure 3.2 and explained in Section 3.2.4. I only assigned a reference point to a particular LULC, if the corresponding LULC class was confirmed in all three layers (S1 FCC, S2 FCC, and high-resolution google earth imagery). Finally, the field data points and points generated through visual interpretation were merged together to be used for training and testing on GEE platform (Figure 2.5a). Representative reference points for monsoon crop and bare soil are shown in Figure 2.5b,c. I randomly identified 70% of the 1500 reference points as 'training points' using the 'randomColumn' function in GEE and used those for training the

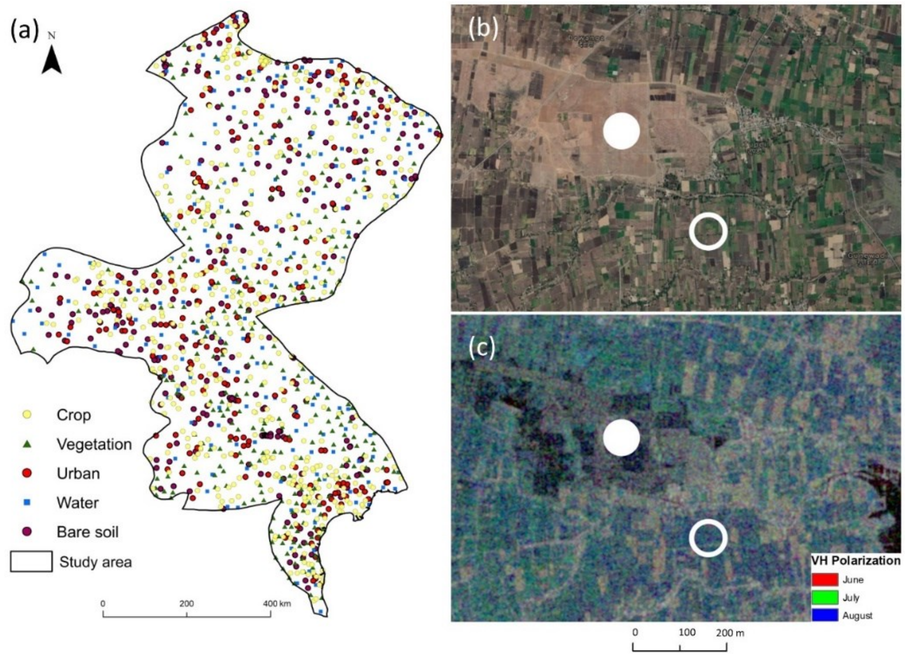

Figure 2.4:(a) Spatial distribution of training and testing points across the agro-ecological regions (AER). The five land use/cover classes used for this study are vegetation (forest/plantation/grass), urban, water, bare soil, and monsoon crop; (b) Representative reference points on the high resolution google earth imagery for monsoon crop (white hollow circle) and bare soil (white solid circle); (c) The same representative reference points as shown in (b) confirmed using Sentinel-1 monthly median false color composite imagery (red—June, green—July, and blue—August).

random forest (RF) classifier. The rest of the reference points (30%) was used as 'testing points', i.e., for post-classification accuracy assessment. To avoid any biases in selecting the training and testing points, I performed the classification and accuracy assessment iteratively for 20 times by randomly dividing training and testing points in 70:30 ratio.

2.2.4 Google Earth Engine platform

Google Earth Engine (GEE) is a cloud-computing platform, which combines open source large peta-byte scale geospatial datasets for planetary scale analysis

(Figure 2.6). It assist the scientists, researchers and developers to detect changes, perform trend analysis and quantify the differences for natural resource management over large area without going into the background of the processing. The open source datasets provided by USGS, ESA, and other organizations are available in ready to use format. It allows scientists to collaborate using datasets, algorithms and visualization tools. It has a repository of vast functions for pre-processing, performing logical and mathematical operations, machine learning algorithms, sampling etc. to perform operations on images. It has the capability of both raster and vector data analysis. The platform uses Python and JavaScript application programming interfaces for making requests to the servers. It also allows users to integrate additional functions using Python and Java script API. Due to its immense capability it has now being used by various organizations such as WRI managing Forest Watch program and integrated in academic curriculum by many educational institutions.

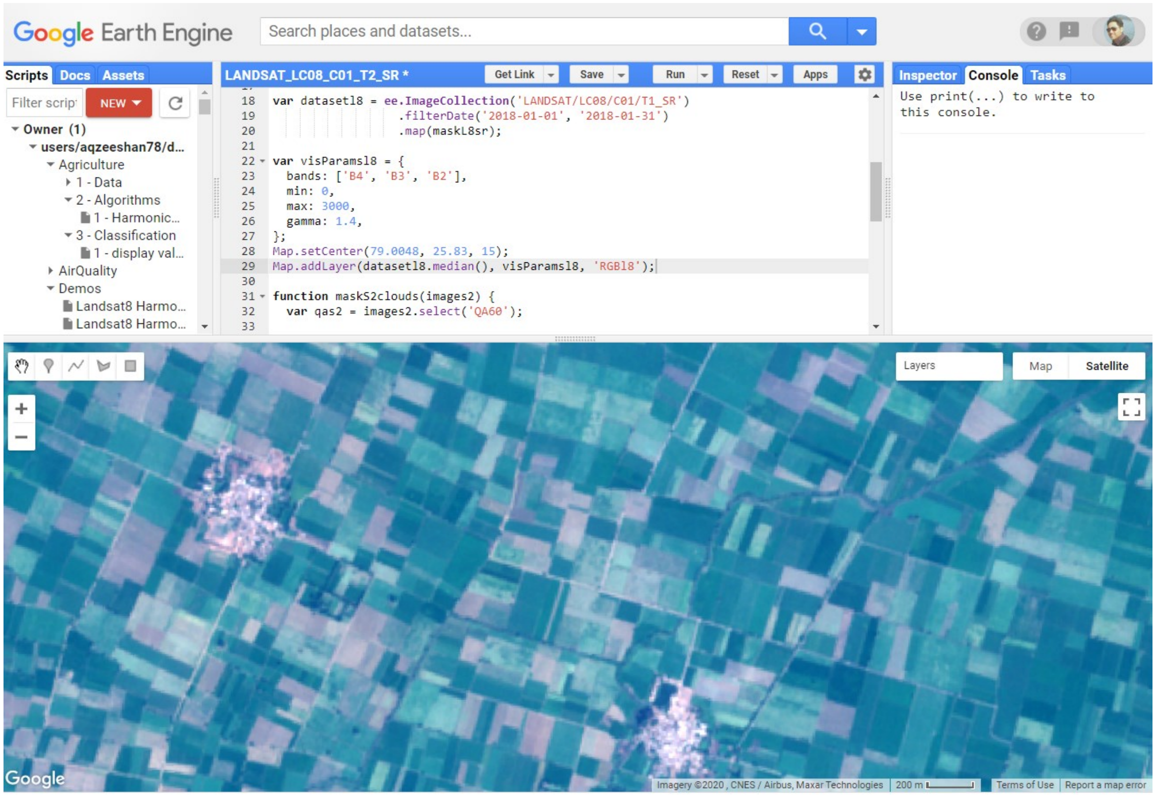

Figure 2.5:The Earth Engine Code Editor at code.earthengine.google.com

Chapter 3

SMONSOON CROPLAND MAPPING USING RADAR DATA

3.1 Introduction

Synthetic aperture Radar (radar) imagining is an active remote sensing which has the capability to collect earth information at all weather capacity (Woodhouse 2006). Radar sensors are influenced by both target and sensor parameters. Radar backscattering are sensitive to the dielectric and geometric characteristics of target features (Sivasankar et al. 2018-1). The radar backscattering signals are also function of sensor parameters such as wavelength of operational, polarization and the angle of incidence (Lone et al. 2017). The different combination of radar sensor parameters produces different results for agriculture monitoring. Hence, appropriate radar sensor parameters are required for to increase the sensor efficiency. radar data helps in distinguishing crops from other land cover classes due to its unique radar backscattering of dynamic agriculture fields. The radar backscattering signal from agriculture fields are combination of crop growth parameters and the underneath soil moisture.

Monsoon cropland, which is the rainfed and major crop-growing season in India, offers more potential to be monitored by radar data. The cropland area, surface roughness and soil moisture are some of the parameters, which can efficiently be extracted using radar data. Using temporal radar data, the change is soil moisture and crop biomass can easily be related to cropland compared to other land cover types. As the crop vegetation grows, the contribution of vegetation in radar backscattering increases and hence are easily observed by radar sensors. The type of crop vegetation, the geometric structure of the plant, height and growth variation and moisture content also have effect on the radar backscattering.

Multi-temporal radar images improve the crop classification accuracy and capture the variation in growth process (Singha et al. 2016, Larrañaga et al. 2016). When the temporal information from radar is combined with multi-polarization then better information of crop conditions are achieved.. Yet, previous studies utilizing radar have been confined to examining croplands dominated by specific water intensive monsoon crops such as rice or jute which are easier to detect due to their distinct backscattering signatures compared to dryland monsoon crops (Sun et al. 2019, Wang et al. 2015). Hence, these radar-based methods need to be evaluated or revised in the context of diverse cropping practices, especially for rainfed monsoon crops grown in dryland regions. Hence, the objective of this study was to Evaluating Sentinel-1 (S1) radar data for monsoon cropland mapping using Machine Learning algorithms.

3.2 Methodology

3.2.1 Overall Workflow

The flowchart for the methods used for this study is outlined in Figure 3.1. In the first step, S1 radar time series images were loaded on Google Earth Engine (GEE) platform using 'ImageCollection' function (Google). These images were then filtered based on time (June-November 2018) and study region boundary. For S1 radar, I used images from June to November. Further image classification was performed on S1 radar (Figure 3.1), using pixel-based machine learning classifier (random forest and SVM) on GEE. I have used pixel-based classifier instead of object-based classifier for large monsoon cropland mapping, as the latter requires high computation time and has complicated intermediate steps including the segmentation where specific parameter tuning is needed (Memarian et al. 2013, Liu et al. 2010). Even though object-based classifiers might improve the classification accuracy in some landscapes, this performance improvement is not always evident in complex heterogeneous landscapes such as the one showed in this study. I further performed accuracy assessments for the four AERs (Table 2.2). I calibrated and validated the algorithms using 1500 reference points collected using high-resolution images. I further re-ran 10 iterations for each algorithm, utilizing unique subsets of the initial training data. Training and testing of the classified images are performed according to the procedure detailed in Figure 2.4 and Section 2.2.3.

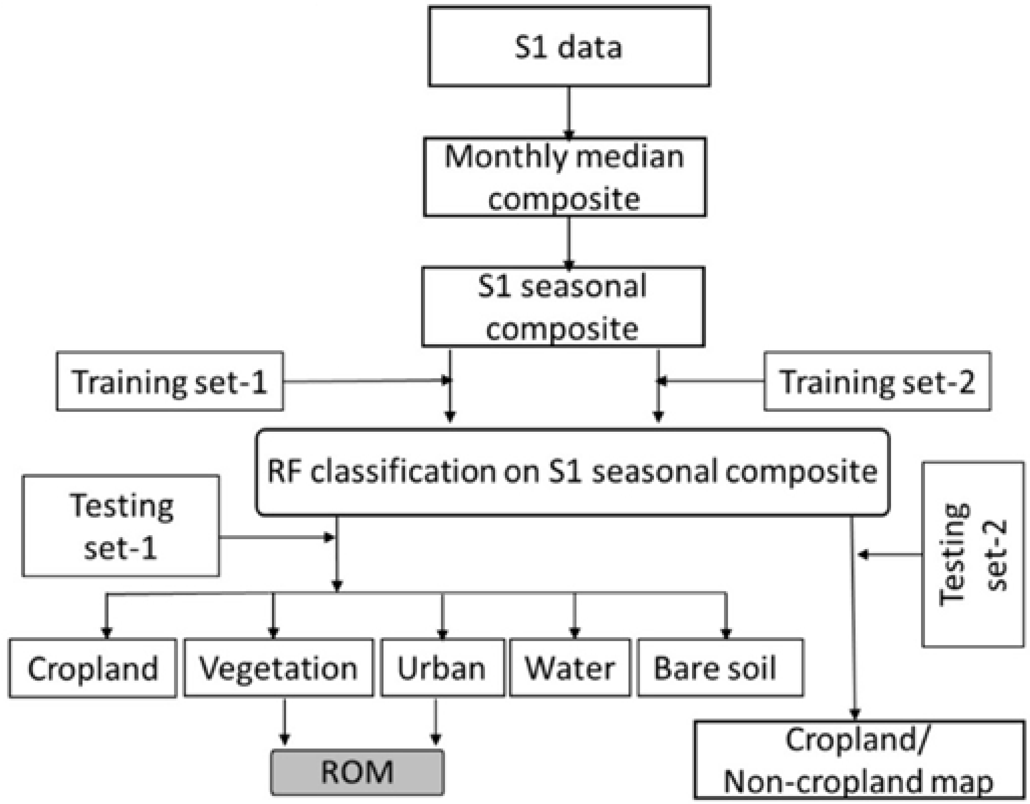

Figure 3.1:Workflow for performing the classification using set-1 reference data to obtain Radar Optical cross Masking (ROM) and using set-2 reference data to obtain crop map for Random Forest (RF) classifier

3.2.2 Accuracy Assessment

Classification outputs obtained from Sentinel-1 radar was evaluated and compared using the standard count-based accuracy assessment methods of overall accuracy (OA) and kappa coefficients obtained from the confusion matrix (Congalton 1991) using 20 different iterations. User's accuracy (UA) and producer's accuracy (PA) were calculated using 30% testing points that were not involved in training the classifiers. UA measures the error of commission, i.e., the proportion of pixels that were incorrectly included in a class that is being evaluated. PA measures the error of omission, i.e., the proportion of pixels in a certain class that is being evaluated that were incorrectly classified in another category, and were omitted from the 'truth' class as identified by the test points. I further calculated the F-score to determine the degree of discrimination among the five LULC classes obtained from the radar-derived classification and the radr based binary crop vs non-crop classification. The F-score ranges between 0 and 1, with higher values denoting better discriminating power among the classes. The F-score is calculated using Equation (1) mentioned below (Powers 2011):

$$\mathrm{F-score} = \left(\frac{(\mathrm{UA} \times \mathrm{PA})}{(\mathrm{UA} + \mathrm{PA})}\right) \times 2 \tag{1}$$

I did not compare the accuracy of the results obtained with the crop estimates provided by government due to non-availability of crop census data for the monsoon crop season 2018–2019.

3.2.3 Satellite Data Pre-preprocessing

The GEE platform provides Sentinel-1 radar data pre-processed with thermal noise removal, radiometric calibration and ortho-rectification using the Sentinel-1 toolbox resulting in ground-range detected images with backscattering coefficients in decibel (dB) scale. Using temporal VH and VV polarization, radar monthly composite images were created by considering median values. I also used these monthly median composite images to create False Color Composite (FCC) to aid in visual interpretation of the images for training and testing data collection as explained previously in Section 2.2.3. The images acquired after the month of November were not considered as I assumed that crops grown after this time are not monsoon crops, based on existing literature (Land Use). A total of 516 Sentinel-1 images and 1734 S2 images were used for the entire monsoon crop-growing season of 2018. I have used multi-temporal stack instead of single images as first, the "median" image shows much lower speckle than the individual images, which improves classification accuracy. Second, different land cover classes show specific behavior over the crop growth period, which are well represented by "standard deviation" of the stack.

3.2.4 Radar Temporal Backscattering

Temporal backscattering profiles were obtained using C-band VH polarization Sentinel-1 imagery from monsoon crops (rice and black gram/soybean), bare soil, urban, water and vegetation (forest/plantation/grass) features shown in Figure 3.2a, similar to what obtained by Singha et al.2019. The backscattering profiles were generated by taking the mean of 10 sample points for each class spread across the study area. The sample points for each class along with their geolocations are shown in Figure 3.2b. Vegetation is defined as land surface with plants and includes plantation, grass and forest. The contrasting nature of backscattering from vegetation and monsoon crops forms the basis of utilizing the temporal Sentinel-1 backscattering signatures for Radar Optical cross Masking (ROM) as backscattering signatures are effective in separating crops (rice, black gram, and soybean) from water, urban and vegetation.

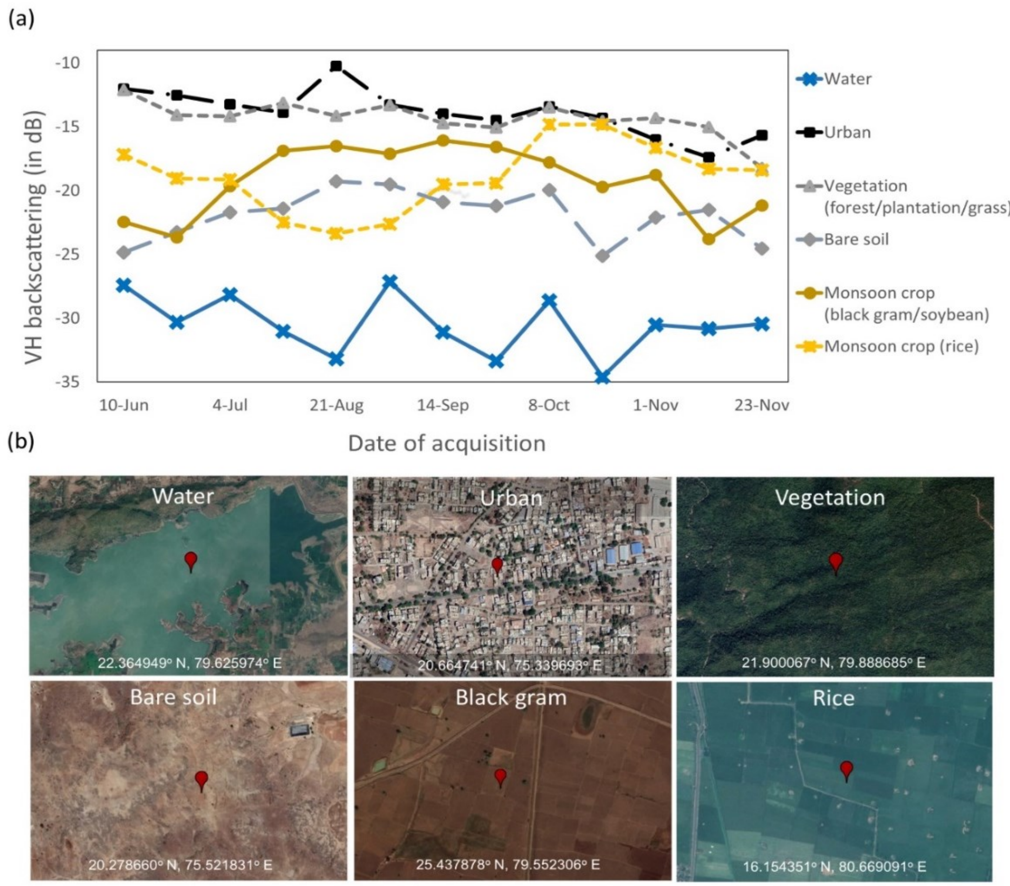

Figure 3.2:(a) Sentinel-1 (S1) mean temporal backscattering profile with VH polarization obtained from 10 points each for land cover features, collected from monsoon crops and other land use/cover classes during the monsoon season (June–November, 2018). Urban and vegetation class shows constantly high backscattering intensities throughout the monsoon season, water shows very low backscattering intensities and monsoon crops and bare soil has backscattering values between urban/vegetation and water; (b) Representative reference points along with its coordinates on the high-resolution google earth imagery.

However, these signatures are mixed with bare soil during the crop-growing season. So, it becomes difficult to segregate crops from bare soil with very high accuracy using only Sentinel-1 radar data. Hindrance in segregating crops from bare soil forms the basis of integrating optical data with radar as bare soil is very distinct in optical data compared to crops and other vegetation due to its lack of 'greenness' reflected in low Normalized Difference Vegetation Index (NDVI) values(Scanlon et al. 2002). In the study region, the radar backscattering signatures obtained from vegetation (forest/plantation/grass) and urban class are non-dynamic throughout the monsoon season, and have nearly constant high backscattering values (~−16 dB to −10 dB) compared to other LULC features. During the time of classification, there is high probability of vegetation class being mixed with urban and vice versa. Monsoon crops and bare soil have dynamic backscattering throughout the crop-growing season. For crops such as black gram and soybean, the land preparation starts from first week of June and last until mid-July based on the onset of the monsoon. For these monsoon crops, the backscattering is initially low due to land preparation in June and increases with time as the crop grows. For rice, land preparation starts in the July/August when the fields have sufficient amount of water as rice is a water intensive crop (Singha et al. 2019). Rice shows very low backscattering during the land preparation/transplanting stage. During the time of maturity, the backscattering increases for both black gram/soybean and rice. The backscattering is high for rice compared to black gram/soybean due to high biomass content resulting in high volume scattering from the rice fields. For bare soil, the initial backscattering is similar to that obtained from black gram/soybean due to the presence of exposed soil with no crop cover. It can be seen that bare soil signature can get mixed with that from rice in the month of July. Hence, overall it is very difficult to segregate monsoon crops from bare soil with very high accuracy. For water, the backscattering is very low (< -25 dB) throughout the monsoon season, hence it is easily segregated from monsoon crops.

3.2.5 Cropland Classification using Sentinel-1 radar data

3.2.5.1 RF based classifier

I considered a monthly composite of radar data using both VH and VV (VH + VV) polarization, instead of a single date image for radar based classification, as previous studies have shown that multi-temporal radar data perform far better than single radar image for crop classification (Clevers et al.1996, Skriver et al. 2011) . Considering multi-temporal radar data becomes even more important for diversified cropping pattern in India as such data are able to take into account the variation of crops grown in different time of the season. Using the training dataset, the RF classifier was run on monthly median composite of June–November, 2018. The RF is an ensemble classifier that follows the decision tree approach in which randomly selected results from multiple decision trees are combined together to obtain highly accurate and stable classification results (Breiman 2001, Tian et aal. 2019). RF algorithm can handle large quantity of complex multi-source data and is robust against overfitting. For initiating RF classifier, the user must define two parameters, the number of trees to grow and the number of variables used to split each node. In this work, the number of decision trees used are 100 and the variables used to split each node was set to square root of the number of overall variables. For this study, radar-based classification was performed using two different output criteria: one with a classified map with five classes and the other with only two classes – cropland and non-cropland (obtained by combining non-cropland classes, i.e., bare soil, water, vegetation and urban) (Figure 3.1). In addition, I calculated the classification accuracy for each AER separately. The classification and accuracy assessments were performed 20 times using unique set of training and testing data.

3.2.5.2 SVM based classifier

SVM is another widely used classifier, which look for optimum hyperplane to separate different classes. The criteria of selecting the support vector depends on the choice of cost parameter C, Gamma and kernel functions. Cost parameter decides the level of punishment for a misclassified data. For SVM classification, the kernel type of linear, polynomial, Radial Basis Function (RBF) and sigmoid function is used. For this study, preliminary analysis was performed using all four kernel type and it was observed that linear function outperforms others and hence linear function was used for further analysis. The gamma function represents the reciprocal of number of classes used. For this analysis, I had two sets, one with five classes and other with two classes and hence Gamma function selected was 0.2 and 0.5 respectively. As for linear kernel, the Gamma value has no significance hence it was dropped for further analysis.

3.3 Results

3.3.1 Radar only classification for five major land cover classes

3.3.1.1 RF based classifier

The classification accuracy obtained from Sentinel-1-derived radar classification using training and testing set-1 for RF is shown in Table 3.1. The OA obtained after 20 iterations is 80.0%. The kappa coefficient obtained is 0.74 (Table 3.1). The results indicate that the water class was easily identified (F-score = 0.96) using radar data among the five LULC classes. The low F-score obtained for urban (0.64) class indicates that the Sentinel-1 radar data has the least discriminative capability to segregate urban from other classes. radar data was moderately successful in discriminating monsoon crops from other land cover classes (F-score = 0.84).

Table 3.1: Accuracy assessment for land cover classes obtained from radar only classification using RF algorithm and VH+VV polarization and training and testing set-1.

Land Cover Type	Radar Only (VH + VV)		
	UA	**PA**	**F-Score**
Water	0.96	0.96	0.96
Bare soil	0.79	0.8	0.79
Urban	0.78	0.54	0.64
Vegetation	0.68	0.75	0.71
Monsoon cropland	0.81	0.87	0.84
OA	0.80		
Kappa	0.74		

3.3.1.2 SVM based classifier

The classification accuracy obtained using SVM classifier along with training and testing set-1 is shown in Table 3.2. The OA obtained after 20 iterations is 77 %. The kappa coefficient obtained is 0.70 (Table 3.2). It can be observed that urban class has the least F-score followed by bare soil. The highest F-score obtained from water body shows that SVM classifier has the highest discriminative capability to segregate water from other classes followed by monsoon cropland. radar data was moderately successful in discriminating vegetation (F-score = 0.75).

Table 3.2:Accuracy assessment for land cover classes obtained from Radar only classification using SVM algorithm and VH + VV polarization and training and testing set-1.

Land Cover Type	**Radar Only (VH + VV)**		
	UA	**PA**	**F-Score**
Water	0.92	0.91	0.92
Bare soil	0.73	0.67	0.70
Urban	0.72	0.54	0.62
Vegetation	0.70	0.81	0.75
Monsoon cropland	0.80	0.84	0.82
OA	0.77 ± 0.011		
Kappa	0.70 ± 0.014		

3.3.2 Radar only classification for cropland mapping using RF and SVM classifier

The comparison of accuracy assessment obtained from radar only using training and testing data set-2 for both RF and SVM are displayed in Table 3.3. The crop vs. non-crop overall classification accuracy obtained by Sentinel-1 satellite data is 90% for RF classifier and 80 % for SVM classifier. The Overall accuracy's standard deviation of kappa values while randomly changing the training and testing points for 20 classifier iterations are equal for both the classifiers. There was a large difference in F-score obtained in cropland discrimination compared to non-cropland while using RF and SVM classifier. The F-score for non-cropland shows high discriminative capability for both RF and SVM classifier compared to cropland mapping.

Table 3.3:Accuracy assessment for crop vs non-crop mapping obtained from Sentinel-1 radar data and training and testing data set-2

Radar Only Classification		**User's Accuracy**	**Producer's Accuracy**	**Overall Accuracy**	**Kappa**	**F-Score**
RF	**cropland**	0.82	0.88	0.90+0.017	0.77+ 0.039	0.85
	non-cropland	0.94	0.91			0.92
SVM	**cropland**	0.74	0.55	0.80 $\pm$ 0.017	0.50 $\pm$ 0.04	0.63
	non-cropland	0.81	0.91			0.86

3.4 Discussions

The low accuracy obtained with the SVM and linear kernel compared to RF can be due to the fact that for this kind of study area having large diversity in agriculture results in diverse training and testing sets which in turn affect the SVM classification accuracy. In addition, the SVM classifiers require the proper optimization of tuning parameters, which becomes difficult to achieve for diversified agriculture systems. In this study, even the other kernels such as polynomial, Gaussian did not produced productive accuracy compared to what I achieved using the linear kernel. The high F-score (>85%) for both cropland and non-cropland shows the high performance of RF based classification on Sentinel-1 dataset compared to SVM classification which has low accuracy in discriminating croplands. In addition, I can observe that RF outperforms SVM in dealing with multi-temporal radar data also. Hence, it can be seen that RF is more robust and less time consuming compared to SVM for monsoon cropland mapping.

3.5 Conclusions

It can be concluded that for monsoon land cover classification and for only monsoon cropland mapping, RF performs better than SVM. Also, both RF and SVM shows almost similar performance for monsoon cropland for User's accuracy and Urban for producer's accuracy.

Chapter 4

MONSOON CROPLAND MAPPING USING RADAR AND OPTICAL DATA

4.1 Introduction

The hindrance in using optical satellite data for intra-seasonal monsoon cropland monitoring over large region requires the remote sensing community to develop new methods, especially for countries with heterogeneous landscapes, such as India. These methods should take into account the variations in cropping practices across different agro-ecological regions (AER). Crop monitoring using optical data has come a long way from the launch of Landsat series satellite data from 1970s onwards. With the advancement of technology and the improvement of spatial, temporal and spectral resolution, the classification results have improved drastically. With the freely availability of Landsat and the launch of Sentinel-2 optical satellite through Copernicus mission by ESA, there is a large utilization of these satellites for crop monitoring. The dependency of optical satellite data on solar energy restricted image acquisition during cloudy conditions. Optical images acquired during this season suffers from cloud cover, haze conditions, and are of limited or of no use. Cloud and cloud shadow remains a major drawback in optical data acquisition and leads to gaps in optical imagery and missing data in time series analysis. Agriculture monitoring in rainfed monsoon cropland are hampered as crop-growing season coincides with the peak monsoon season and hence image acquisition and image classification is hampered as there are no or very limited in-season images available for monsoon cropland mapping over large regions.

To overcome the cloud and haze issues, the multi-sensor combination offers one of the solutions by exploiting different parts of electromagnetic spectrum which are not affected by clouds. Microwave radiations offers one of the solution as these radiations are not attenuated by cloud cover (Woodhouse 2006, Sivasankar et al. 2018-2, 2019). Satellites sensors exploiting the use of electromagnetic radiation in the microwave portion and sending their own energy pulse and measuring the reflected pulse from the target on the ground are known as radar. radar uses the motion of the instrument for image acquisition at satisfactory resolution (Woodhouse 2006). Owing to the difference in image acquisition, satellite data from radar and optical imagery are found to be complementary. The synergy of radar and optical sensor open new arenas for the development of new classification methodology that can exploit the advantages of both sensors. Previous studies across the world has shown the importance of combining both the sensors data to improve the cropland mapping, crop type identification and classification. Studies have reported an increase of atleast 5% improvement from using optical alone and combining optical with radar imagery. However, despite the increasing efforts to combine optical and radar imagery, very few studies has been done to monitor monsoon cropland over large region using combination of optical and radar imagery. There are very few studies, which combine radar and optical data for monsoon crop mapping, but these studies are restricted both spatially and temporally to be used only for a small region. These studies either focused on water intensive monsoon crops or over small regions where obtaining a few optical image snapshots was possible during the monsoon season (Verma et al. 2019, Kumari et al 2019). However, there are no studies performed by using radar data alone or by integrating radar and optical data for extracting monsoon cropland over a large area in different agro-ecological regions practicing diverse agriculture systems. This study intends to fill the gap in monsoon cropland monitoring by combining radar and optical data and evaluate the combination of radar and Sentinel-2 (S2) optical data in terms of providing greater accuracy for monsoon cropland mapping.

4.2 Methodology

4.2.1 Overall workflow

The flowchart for the methods used in this study is outlined in Figure 4.1. In the first step, Sentinel-1 (S1) radar and Sentinel-2 (S2) optical time series images were loaded on Google Earth Engine (GEE) platform using 'ImageCollection' function (Google). These images were then filtered based on time (June-November 2018) and study region boundary. For Sradar, I used images from June to November, but for S2 optical data, I considered images from July to November. The month of June was not considered for S2 optical data as summer crops were still at their peak growth stage in some regions and the land

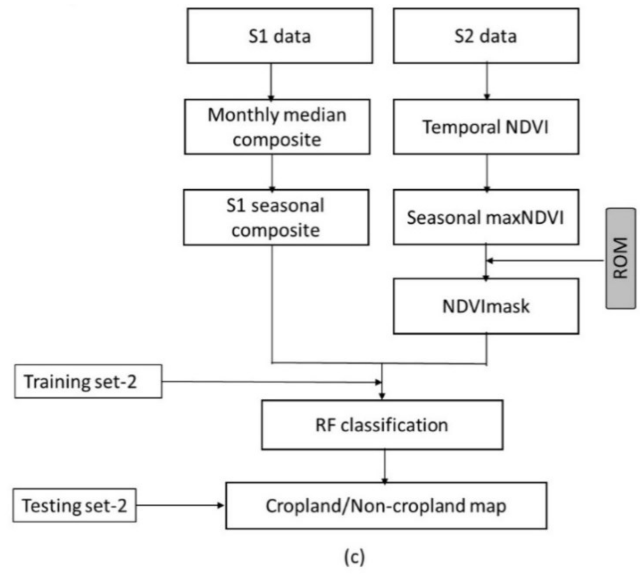

Figure 4.1:Overall workflow followed For radar (S1)+optical (S2) combined classification using set-2 reference data using the Random Forest classifier.

preparation and sowing of monsoon crops were in the initial stages. A ’cropped field’ in June would thus be an indication of summer crops, and not monsoon crops. Further image classification was performed on radar+optical combined (Figure 4.1), using pixel-based machine learning classifier (RF and SVM) on GEE. I have used pixel-based classifier instead of object-based classifier for large monsoon cropland mapping, as the later requires high computation time and has complicated intermediate steps including the segmentation where specific parameter tuning is needed (Liu and Xia, 2010). Even though object-based classifiers might improve the classification accuracy in some landscapes, this performance improvement is not always evident in complex heterogeneous landscapes such as the one showed in this study. I further performed accuracy assessments for the four AERs (Table 2.2). Training and testing of the classified images were performed according to the procedure detailed in Figure 2.4 and Section 2.2.3. I calibrated and validated the algorithms using 1500 reference points collected using high-resolution images. I further re-ran 20 iterations for each algorithm (RF and SVM), utilizing unique subsets of the initial training and testing datasets.

4.2.2 Accuracy Assessment

Classification outputs obtained from S1 radar data was evaluated and compared using the standard count-based accuracy assessment methods of overall accuracy (OA) and kappa coefficients obtained from the confusion matrix using 20 different iterations. User's accuracy (UA) and producer's accuracy (PA) were calculated using 30% testing points that were not involved in training the classifiers. UA measures the error of commission, i.e., the proportion of pixels that were incorrectly included in a class that is being evaluated. PA measures the error of omission, i.e., the proportion of pixels in a certain class that is being evaluated that were incorrectly classified in another category, and were omitted from the 'truth' class as identified by the test points.

I further calculated the F-score to determine the degree of discrimination among the five LULC classes obtained from the S1-derived classification and the S1 based binary crop vs non-crop classification. The F-score ranges between 0 and 1, with higher values denoting better discriminating power among the classes. The

F-score is calculated using Equation (1) I did not compare the accuracy of the results obtained with the crop estimates provided by government due to non-availability of crop census data for the monsoon crop season 2018–2019.

4.2.3 Pre-processing of Sentinel-2 data

Sentinel-2, Multi-spectral level 1-C processing provided by GEE were used (Sentinel 2 Handbook). These data have been ortho-rectified and radio-metrically corrected atmosphere (TOA) reflectance data product. The radiometric and geometric corrections were performed according to the method demonstrated in the Sentinel-2 User Handbook (Sentinel-2). I used the GEE function “ee.ImageCollection” to filter the time series imagery according to the required dates and the selected study region. As we were interested to extract only the variation in Normalized Difference Vegetation Index (NDVI) in the region, I selected only the 10m spatial resolution, Red and NIR band for this study. The TOA reflectance data product generally contains considerable atmospheric signals, hence the data has to be corrected for atmospheric signals. Automatic cloud and cloud shadow masking algorithm was used for reducing the atmospheric effects on the reflectance data. Also, I considered the images having cloud cover of 5% or less to reduce the atmospheric affects further. For performing temporal aggregation to generate max NDVI, each pixels will be containing varying number of images based on cloud free image availability. For S2 optical data, I considered images from July to November. The month of June was not considered for S2 optical data as summer crops were still at their peak growth stage in some regions and the land preparation and sowing of monsoon crops were in the initial stages. The images acquired after the month of November were not considered as I assumed that crops grown after this time are not monsoon crops, based on existing literature (NSFM). A total of 1734 S2 images were used for the entire monsoon crop-growing season of 2018.

4.2.4 Seasonal Normalized Difference Vegetation Index (NDVI)

The Normalized Difference Vegetation Index (NDVI) is a widely used remote sensing measure to assess the health of the vegetation and to differentiate crops, and other vegetation (forest, plantation, grass) from bare soil, water and urban (Tucker 1979). NDVI is a unit less measure and ranges between -1 and 1. Healthy vegetation typically has higher NDVI values compared to non-vegetated surfaces. For calculating NDVI, I require the red and near-infrared (NIR) reflectance bands (Equation 2):

$$NDVI = (NIR-Red)/(NIR+Red) \tag{2}$$

In S2 imagery, I used band 4 and band 8, respectively, for red and NIR in the above equation. I calculated NDVI for all the available cloud free pixels in the image as main focus was on cropped field identification, especially since previous studies have shown high correlation between NDVI and photosynthetic activities of the cropped fields (Benedetti and Rossini 1993). To generate the seasonal maximum value of NDVI (maxNDVI), I performed temporal aggregation of NDVI from July to November (Figure 4.2). Temporal aggregation is an approach to perform pixel-based analysis over a period of time using metrics (i.e., mean, median, maximum etc.) from satellite derived reflectance or satellite-derived indices (e.g. NDVI, Enhanced

Vegetation Index (EVI), etc.)(Carrasco et al 2019). The aggregation addresses the problem of lack of continuity in the optical data due to cloud cover and reduces the volume of data for further processing (Pericak et al. 2018). During monsoon season, optical satellite images in India contain considerable amount of cloud patches, which affects the radiometric quality of the images, thus limiting intra-seasonal crop monitoring capability. I calculated maxNDVI in order to fill this data gap and to capture the crop heterogeneity, i.e., considering all monsoon crops with different intra-seasonal phenology.

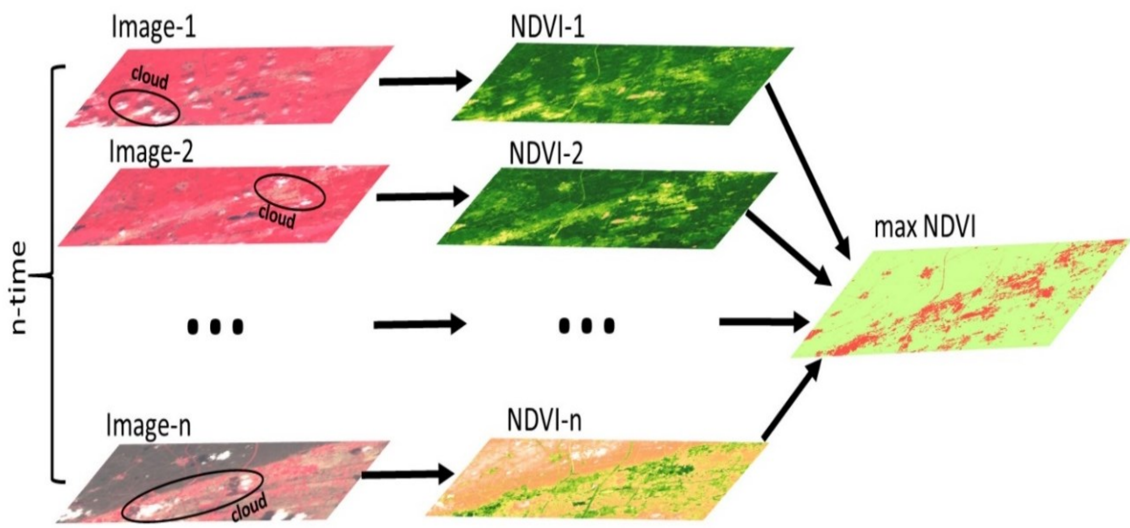

Figure 4.2: Temporal aggregation of normalized difference vegetation index (NDVI) derived from seasonal sentinel-2(S2) data to obtain the maxNDVI during the monsoon season.

4.2.5 Otsu thresholding

MaxNDVI generated using S2 NDVI contains both vegetation and non-vegetation features such as Urban, Water, bare soil which has very low pixel DN values compared to vegetative maxNDVI. The segregate vegetation from non-vegetation both supervised and automatic thresholding based techniques can be used. However, using supervised based techniques offer another challenge as it requires training data before hand to run the classification. Hence, in such situation many times automatic methods are preferred. Unsupervised classification techniques and other decision rule based techniques are preferred, as they do not require any training points. Otsu thresholding used in this work is one of the method, which is clustering based thresholding method to automatically find optimum thresholding to separate two different kind of pixels, by performing the histogram analysis on the observed pixel values. Using Otsu thresholding, I can easily segregate two types of relatively homogenous land cover features such as water vs land, vegetation vs land etc. Otsu method assumes that the two classes are separated by a bimodal histogram of the digital values. Otsu thresholding had gained popularity especially in image processing techniques and in satellite image analysis for detecting flooded areas in radar imagery as radar imagery offers a better potential for discriminating water features from other land cover features due to low backscattering from water bodies in the radar imagery. In this maxNDVI dataset, both crops and vegetation (forest/plantation) have higher values compared to water, urban and bare soil. I further utilized the Otsu's thresholding approach in GEE, to differentiate between the crops/vegetation (forest, plantation) from non-vegetative features with low NDVI values (bare soil, urban and water) (Liu et al. 2002). This approach is an automated way of finding an optimal global threshold based on the observed distribution of pixel values. Based on the pixel value distributions for the LULC classes considered in this study, the Otsu's thresholding value obtained for maxNDVI is 0.36.

4.2.6 Radar Optical cross Masking (ROM)

It was not possible to differentiate between crop and vegetation within the maxNDVI dataset, which is an important step for a crop mapping procedure involving optical data. Due to non-availability of high-resolution (10m) non-crop mask for the region, I developed a method of masking non-crop vegetation from maxNDVI using classified radar imagery (Figure 4.1). From the S1 radar-derived classified map with five LULC classes, vegetation (forest/plantation/grass) and urban classes were further combined together as a non-dynamic class to obtain the non-crop mask and to segregate crops from vegetation in maxNDVI imagery. I combined vegetation and urban as one class, instead of considering only vegetation, since these two classes have similar backscattering signatures and are difficult to segregate in S1 radar-derived map as described previously (Section 3.2.4, Figure 3.2). Moreover, combining urban and vegetation to obtain the non-crop mask is less likely to affect the outputs, as urban class is already masked out from the maxNDVI data due to the application of Otsu's thresholding. I coined this technique as Radar Optical cross Masking (ROM) where I used the non-dynamic, non-crop (urban + vegetation) mask to separate vegetation from crops resulting in crop only maxNDVI dataset (NDVImask; Figure 4.1)

4.2.7 Classification based on combined radar and optical data

In this step, the S1 radar data and NDVImask (optical) were combined for pixel-based classification to examine if adding NDVImask imagery will result in improved monsoon crop mapping accuracy compared to using radar only classified map. Combining S1 and NDVImask will also address some of the limitations of using only radar data for monsoon crop mapping (Section 3.2.4, Figure 3.2). The monthly median radar composites from June to November were stacked together with NDVImask data obtained after using ROM. The RF classifier was run on the combined dataset with the number of trees set as 100 and the variables to split each node set to square root of the number of overall variables. The output from this classification is a binary crop/non-crop map using training and testing set-2 (Figure 4.1). Similar to the radar-based classification, the combined radar and optical-based classification and accuracy assessment were repeated for 20 times to avoid any biases in the classification accuracy.

4.3 Results & Discussions

4.3.1 Cropland mapping based on combined radar and optical data

The per class accuracies both producer's and user's accuracy and kappa coefficient obtained from radar+optical (S1+S2) combination using training and testing data set-2 is displayed in Table 4.1 using the RF and SVM based classifier. The crop vs. non-crop overall classification accuracy obtained by radar + optical combination is approximately 3% higher than obtained by using radar only dataset

(Table 4.1). The standard deviation of kappa values while randomly changing the training and testing points for 20 classifier iterations are slightly higher for radar only compared to radar + optical classification. The F-score shows high discriminative capability (> 0.85). Moreover, the F-score for the combined radar + optical is higher compared to radar only classification for both crop and non-crop class. Also, the classification accuracy obtained for the combined approach is very high (~10%) in SVM compared to using only radar only data. While both the classifiers show high performance

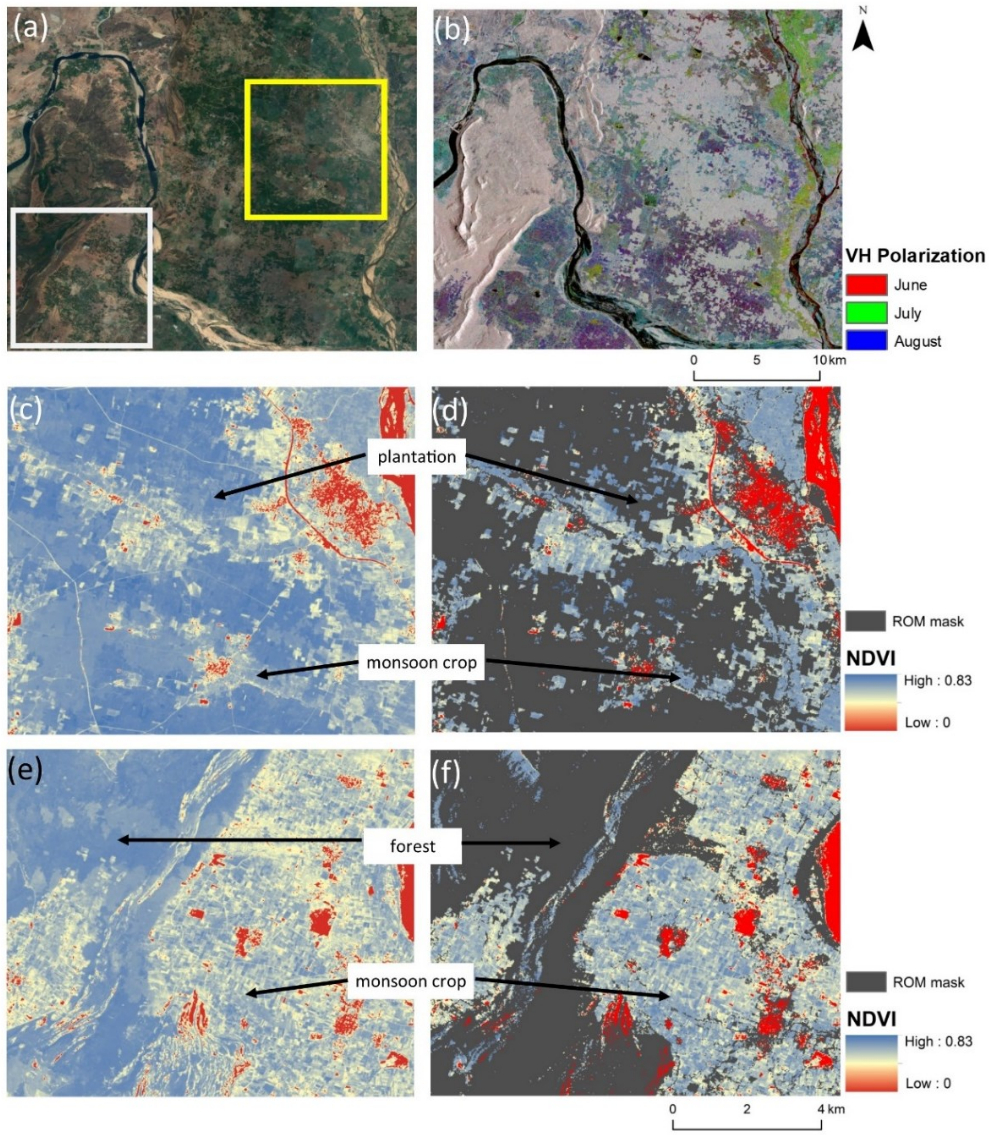

Figure 4.3:Steps for obtaining high-resolution (10m) non-crop mask using the ROM technique (a) High-resolution google earth imagery showing forest class mixed with monsoon crops in white square box and plantation mixed with monsoon crops in yellow square box; (b) False Color Composite VH polarization Sentinel 1 (S1) radar imagery for the same region; (c) maxNDVI for plantation region before applying ROM; (d) NDVImask obtained after applying ROM for plantation; the plantation regions are masked out from monsoon crop and is shown in the dark grey color; (e) maxNDVI for forest region before applying ROM; and (f) NDVImask obtained after applying ROM for forest region. It can be observed that regions of hill shadows are not masked completely.

(>90%), the resulting variations in accuracy (standard deviation) shows more variations in RF compared to SVM whereas it was not the case for SVM classifier for Sentinel-1 datasets. This may be due to high performance of SVM classifiers for optical datasets as compared to radar datasets.

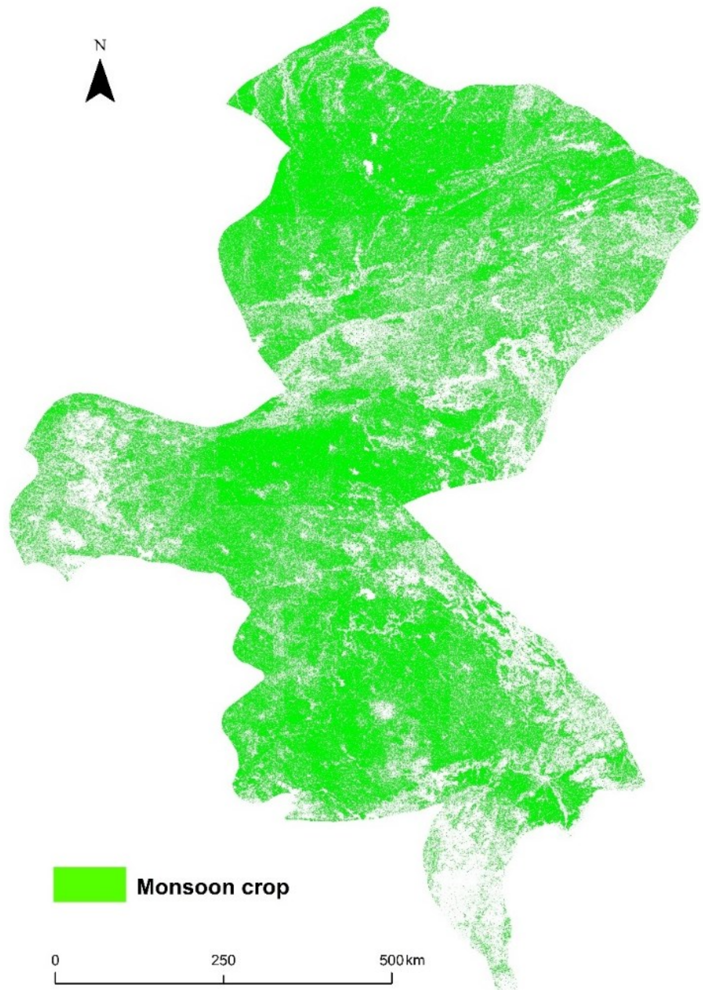

Figure 4.4:Monsoon cropland map obtained using radar+optical (S1+S2) combination and training and testing set-2.

NDVImask image obtained after applying ROM on maxNDVI is shown in Figure 4.3. It can be seen that non-dynamic forest/plantation regions can be effectively separated from crops using ROM. In the figure, regions with forest and plantation (casuarina and eucalyptus) are masked out using ROM to obtain NDVImask image with crops only. The crop map generated using combined radar+optical data is shown in Figure 4.4. Detailed zoom-in views for selected locations using the combined radar+optical classification and its comparison with high-resolution imagery are also shown (Figure 4.5). It can be observed that the combined approach is efficient in differentiating monsoon cropland from plantation (such as mentha/casuarina/eucalyptus) in AER-1 (Figure 4.5a) and AER-4 (Figure 4.5c).

4.3.1.1 Accuracy of combined radar and optical data based cropland map using RF and SVM based based classifier

Table 4.1:Accuracy assessment for crop vs non-crop mapping obtained from radar and optical data and training and testing data set-2 for Random Forest (RF) and Support Vector Machine (SVM) classifier

Radar+Optical Classification		**User's Accuracy**	**Producer's Accuracy**	**Overall Accuracy**	**Kappa**	**F-Score**
RF	**cropland**	0.88	0.9	0.93+0.015	0.83+ 0.033	0.89
	non-cropland	0.95	0.94			0.95
SVM	**cropland**	0.81	0.88	0.90 ± 0.012	0.78 ± 0.026	0.84
	non-cropland	0.94	0.91			0.92

4.3.2 Accuracy of Binary Crop Maps for each AER

The comparison of accuracy assessments obtained from radar only and radar + optical combination using training and testing data set-2 for the selected AER regions are shown in the Table 4.2 & 4.3. I found that for all of the AERs, OA obtained from radar + optical combination outperformed the one obtained from radar only classification and the improvement varies across the AERs. Also, the OA obtained by the combined radar + optical was greater than 90% for all of the AERs. The OA difference between radar + optical and radar only was the lowest for the AER-3 (Table 4.2 & 4.3), whereas it was greater than 4% for the other three selected AERs. For AER-1, which is dominated by rainfed-dryland crops (90%) with some rice-growing regions (10%), there is a 4% improvement in classification accuracy from radar to radar + optical.

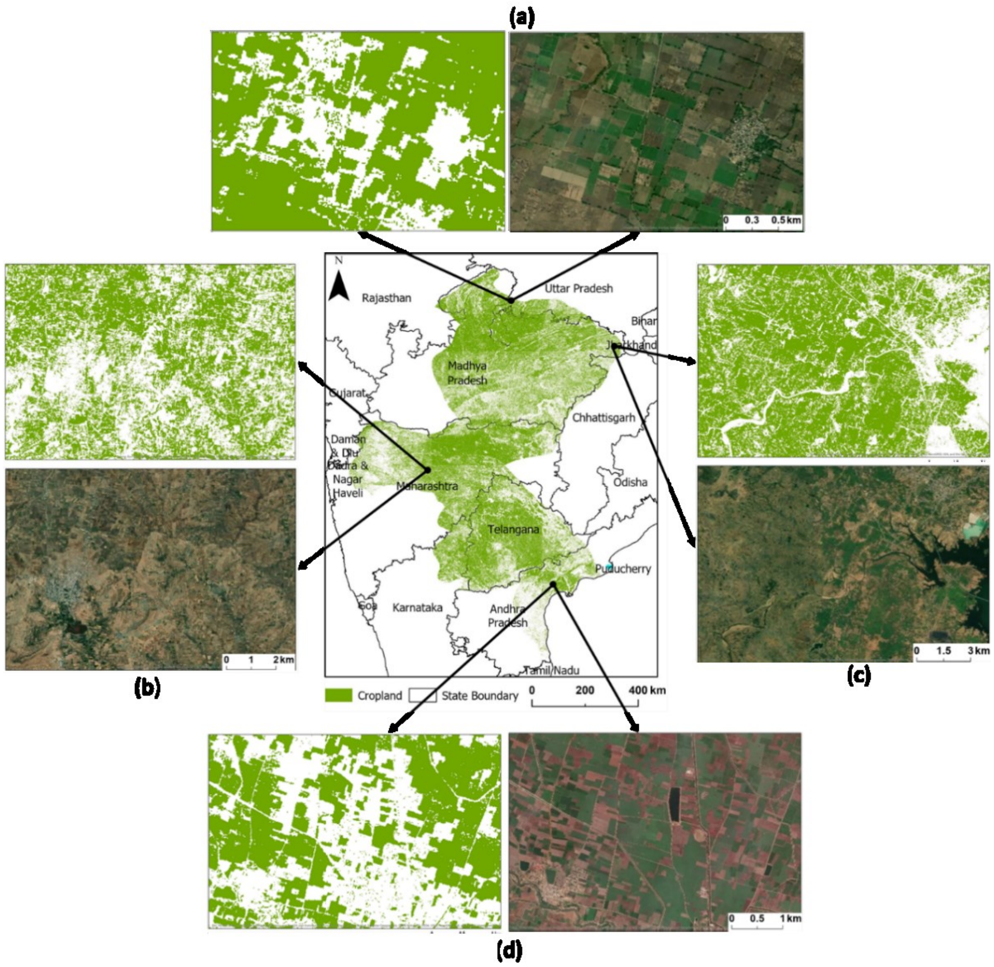

Figure 4.5:Zoom-in view of the monsoon cropland map generated from the combination of radar+optical (S1+S2) for the agro-ecological regions (AER) at various scales and its comparison with high resolution imagery: (a) Northern Plain (AER-1); (b) Deccan plateau (AER-3); (c) Central Highlands (AER-2); and (d) Deccan Plateau, Eastern Ghats and Eastern coastal plains (AER-4 and 5).

For the S1 radar dataset, low classification accuracy of AER-2 and AER-4 and 5 compared to the other two AERs is due to the fact that these two regions are dominated by vegetation mixed with crops and have hilly undulating terrain which may have reduced the radar only classification accuracy. AER-2 hosts Vindhyachal and Satpura range whereas AER-4 and 5 are dominated by Eastern Ghats and fragmented vegetation. For both these regions, classification accuracy improved by 5% when radar data is combined with optical data. In addition, late maturity crops such as rice or cotton dominate these two regions. Hence, contribution from S2 data towards classification accuracy increases in these regions with increasing availability of cloud free optical data towards the end of the monsoon season. For AER-3, it was observed that combined radar + optical dataset shows no major improvement over radar only classification. This region is mainly a plateau with less variation in elevation and negligible forested land. Hence, the mixing of crops with natural vegetation is limited resulting in no major inaccuracy in radar-derived classification due to terrain or vegetation.

Table 4.2:Classification accuracy for different AERs obtained from S1 radar and combined radar+optical classified maps using RF classifier.

Radar only Classification				**Radar+Optical Classification**		
	OA	**Kappa**			**OA**	**Kappa**
AER-1	0.90	0.81		AER-1	0.94	0.88
AER-2	0.89	0.76		AER-2	0.94	0.86
AER-3	0.92	0.79		AER-3	0.93	0.83
AER-4 and 5	0.85	0.67		AER-4 and 5	0.90	0.77

Table 4.3:Classification accuracy for different AERs obtained from radar only and combined radar+optical classified maps using SVM classifier

Radar only Classification				**Radar+Optical Classification**		
	OA	**Kappa**			**OA**	**Kappa**
AER-1	0.84	0.68		**AER-1**	0.88	0.76
AER-2	0.79	0.47		**AER-2**	0.89	0.74
AER-3	0.84	0.58		**AER-3**	0.91	0.78

AER-4 and 5	0.79	0.50		**AER-4 and 5**	0.90	0.78

4.4 Conclusions

Overall, it can be observed that RF outperforms SVM in the combined approach of using radar+optical dataset. However, it was also observed that the performance of SVM increases at higher rate compared to RF in the combined approach. In addition, the SVM shows more robustness and fewer variations in the accuracy using the combined approach compared to the RF classifier. It can be inferred that for monsoon cropland for smallholder farmers study RF outperforms SVM not only for the overall study but also even for each AER of the study.

Chapter 5

DISCUSSIONS

5.1 Monsoon Crop Mapping by Combining radar and optical data

Due to the lack of free time-series of radar data until recently, previous studies mostly focused on using medium resolution Landsat data (30m) or MODIS data (250 m) over large geographic region for monsoon cropland mapping (Pittman et al. 2010, Granados Ramirez et al. 2004). However, using MODIS or Landsat might not be the best approach for monsoon cropland mapping due to frequent cloud cover (Jain et al. 2013, Whitcraft et al. 2015). The spatial resolution of these coarse resolution satellites is not suitable either to capture the small field sizes or mixed agriculture landscapes, thus limiting their usage for preliminary assessment and understanding of croplands over large region. Relying on cloud free optical data alone is not always viable for studying monsoon crops as most of the crops are harvested before cloud free scenes become available in the late monsoon season. Using radar data during the monsoon season can address this issue. However, radar data suffers from speckle effects, which makes it difficult to use radar data alone for generating reliable crop statistics across large regions (Tian et al. 2019). Both optical and radar sensors have limitations for monsoon crop study, but a synergistic approach of combining these data can improve the crop mapping for small-scale farmers at high resolution (Figures 4.4 and 4.5). The technique used in this study differ from other published literature as I propose a new way of pixel-based combination of radar data with temporal aggregation of optical data (maxNDVI) using ROM. Previous studies for monsoon crop monitoring using combination of radar and optical data were limited to water-intensive rice crops and/or small geographic regions where it was possible to obtain at least one cloud free optical image. The results presented in this work are important, as this will provide the first high-resolution (10 m) monsoon cropland map generation, and can also be transferred to other agro-ecoregions. This method shows an improvement over existing methods that are primarily focused on non-monsoon/winter cropland mapping at 30 m or coarse resolution (Becker-Reshef et al. 2010).

The ROM generated from radar data, addresses the issue of miss-classification of spectrally similar plantation and forested vegetation with monsoon crops as visually interpreted (Figure 4.3). During the monsoon season, optical datasets are only available towards the end of season when crops have already been harvested or are in their peak growth stages (Whitcraft et al. 2015). During the peak-growth stage, the spectral signatures of these crops are similar to plantation or other non-crop vegetation, thus making it difficult to segregate the monsoon crops from natural vegetation (Singha et al. 2019, Mercier et al. 2019). The usage of temporal radar-based phenology to generate five land cover classes to produce ROM, masks out the vegetation from monsoon cropland, and improves the classification accuracy (Figures 4.3 and 4.5). ROM helps in segregating monsoon cropland from plantation and natural vegetation (forest/grassland) and can be utilized for large regions, as it is not affected by clouds. The ROM produced here is dynamic and can be regularly updated based on the available radar images. The ROM may also have applications in LULC change monitoring and segregation of non-dynamic LULC features from dynamic croplands.

Overall, this method of integrating radar composite with seasonal NDVImask for monsoon cropland mapping overcomes four main challenges of mapping smallholder agriculture across large spatio-temporal scales: (i) the method works well in different agro-ecological regions as it takes into consideration of the crop planting time and duration, (ii) it can be used in regions with high cloud cover, such as most tropical countries, (iii) it reduces the sub-pixel heterogeneity in mapping monsoon cropland as the resolution of the output cropland map (10 m) better matches the small farm/cropland sizes in most developing countries, and (iv) it helps in distinguishing between monsoon cropland areas from plantation/natural vegetation which has similar signatures during the peak crop growing season. The high-resolution monsoon cropland map produced in this work has the potential to assist government agencies, landscape managers, and researchers in monitoring monsoon crops, which in turn would help us to better understand the factors influencing the production of these crops. Currently, it takes more than a year to make these crop estimates available for decision makers and researchers. This study also has the potential to support global agriculture monitoring missions of Sen2Agri and Geo Global Agriculture Monitoring (GEOGLAM). The objective of GEOGLAM is to provide timely, easily accessible scientifically validated remotely sensed data and derived products for crop-condition monitoring and production assessment. Also, one of the requirements of Sen2Agri mission is to produce national scale dynamic cropland masks other than producing cloud free composites, crop type maps and to indicate the status of current vegetation at 10 m resolution (Becker-Reshef et al. 2010, Defourny et al.2019) Previous research conducted for Sen2Agri mission to generate dynamic cropland was limited in scope in tropical regions as they relied only on optical datasets (Defourny et al.2019, Inglada et al. 2015). This work supports the GEOGLAM and Sen2Agri mission as it produces high-resolution monsoon cropland map over large region comprised of different crop-growing regions. The methodology developed here is also suitable for generating national level dynamic cropland masks.

5.2 Performance of Machine Learning classifier

The classification results obtained by using the machine learning classifiers are impacted by many factors. One of the factors is the accuracy of training and validation dataset used in the study. Same classifier may produce different results on varying the training and testing datasets (Mondal et al. 2019). Hence, the cross-validation technique used here by using 20 different sets of training and validation datasets and averaging them together to generate the results is robust. This also avoids inconsistency in classification and false positives. It was also observed that overall RF outperformed SVM for this study. One of the reasons may be that RF can handle large quantity of complex multi-source data from combination of radar and optical data compared to SVM. RF can handle large database of temporal images and requires less training time. In addition, the number of user-defined parameters required in RF is less and easier to define compared to SVM (Kamusoko et al. 2014, Toosi et al. 2019). In addition, RF performs better if there are sufficient amount of training dataset similar to what I have in this study. Previous studies have also shown that the accuracy of RF increases compared to SVM as the training sample increases and SVM shows better accuracy in limited training dataset. This may be due to the underlying behavior of SVM as it uses smaller subset of training sets even if the training sample is high. SVM is more sensitive to the choice of training data as compared to RF. In addition, SVM are more sensitive to the choice of Cost and kernel parameters and in turn effect the classification results. For this study, linear kernel was used on complex radar dataset hence the SVM classification accuracy may be reduced. Further tuning of cost parameters accompanied by varying kernel might help achieve better results for SVM in higher sample size.

5.3 Error source in the results

Several sources of error might have affected the results from the radar+optical combined methodology presented in this study. One of the reasons may be due to the lack of cloud free S2 pixels during the crop growing season. It is possible that in some regions the classification results were solely generated from radar data due to the non-availability of a single cloud-free S2 image, and could result in inconsistencies in accuracy. The quality of training and validation samples may also affect the classification results. As majority of the training and validation samples for this study was chosen using the satellite imagery without ground verification over large region, this might introduce a possible source of error in the classification results. There might be errors while training the model and/or due to mixed pixels which may reduce the classification accuracy. The study area is complex with varying farming practices which may result in misclassification of land cover classes in radar data. Also, the variations in radar backscattering due to geometric errors (layover, shadow) over hilly terrain affects the accuracy (Singha et al. 2019). Using temporal radar data, along with stratified random sampling and running multiple iterations of the RF classifier reduces the biases, however, does not completely eliminate it. Using the automatic Otsu thresholding method to extract vegetation cover and segregate vegetation from low NDVI values representing soil, water, and other non-vegetated regions in optical data reduces the overall uncertainty as well.

5.3.1 Pixel Level Agreement between the RF and SVM classifiers

The Pixel Level Agreement (PLA) between the RF and SVM shows that both the classifier performs with high accuracy in the region On visual inspection, it was observed that the main difference between the two classifier primarily occurs in the region where there is close juxtaposition from cropland to non-cropland or vice-versa. For example, in the Figure 5.1 shown below, the Pixel level dis-agreement is dominated in the region where is transition happening from dense urban to cropland in AER-4 showing Guntur city. It may be because while RF is more robust in identifying both monsoon cropland and non-cropland, SVM is not comparably effective (Table 4.1, 4.2). RF is also successful in discriminating classes with similar characteristics such as natural vegetation (forest) and cropland during the peak growth stage (Akar and Gungor 2012). Due to the nature of SVM, it has more affinity to boundary pixels and hence there are more issues in SVM data in the transition or boundary pixels from cropland to non-cropland or vice versa. The PLA map generated for the two classifier gives us the confidence in the results and shows us the region where there is disagreement and hence more focus has to be given while implementing the results on the ground.

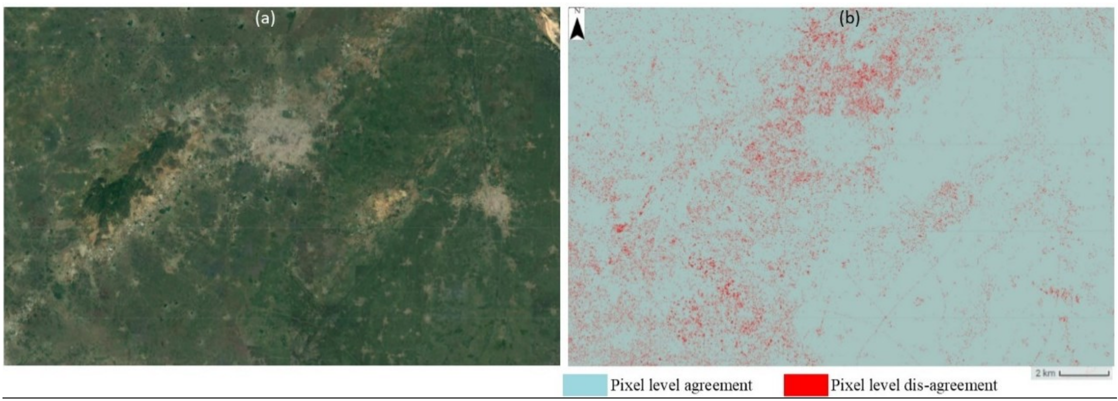

Figure 5.1:Pixel Level Agreement map for the RF and SVM classifier (b) in comparison to the Google Earth imagery for part of AER-4 region (a).

5.3.2 ROM Uncertainty

In this study, radar-based classification was performed using the RF classifier and training and testing data set-1 for generating ROM. The accuracy of ROM and in turn NDVImask depends on how accurately the non-dynamic land use/cover classes are classified. Based on the classification accuracy (Table 3.1, 3.2), it was observed that the producer's accuracy (PA) was the lowest for the urban class. There were many instances where urban area on the ground was misclassified as other classes including vegetation, likely due to the presence of tree canopy cover in urban centers. The accuracy of ROM will vary depending on whether these omitted ‘urban’ points are being classified as ‘vegetation’ or other classes. The results indicate that the user’s accuracy (UA) was the lowest for vegetation class (Table 3.1, 3.2) which shows that points from other classes were committed to the vegetation class. The overall accuracy of the classification will also affect the performance of ROM. The F-scores for urban (0.64) and vegetation (0.71) show low discriminative capability compared to the water (0.96) and monsoon crop (0.84) classes (Table 3.1, 3.2). Thus, this may also have affected the accuracy of ROM. Visual inspection of the output maps revealed that the classification accuracy of radar data to obtain ROM was high for plantation compared to forested regions (Figure 4.3 c–f). This is due to the fact that in the study region, forested regions are mainly found in hilly and mountainous regions, which are affected by geometric errors such as layover or shadow and thus affect the classification accuracy. Also, the forested regions in this part of India is either open forest or scrubland which has open spaces or bare soil in between the canopies, affecting the accuracy of ROM (Roy et al. 1996, Mondal et al. 2020).

To improve the classification accuracy of radar data for ROM generation, second order texture measures, which involves using Grey-Level Co-occurrence Matrices (GLCM), can be included for improving the classification accuracy of radar data, especially for discriminating forest and plantation regions. In addition, with advancement in technology and availability of large amount of satellite data, more powerful deep learning methods such as long short-term memory (LSTM), which efficiently handles time series data, may be utilized for improving the overall classification accuracy and in producing ROM in particular (Massey et al. 2018, Rubwurn and Korner 2017).

Chapter 6

CONCLUSIONS

This study presents a synergistic approach of combining radar with optical data for monsoon cropland mapping over different agro-ecological regions in India utilizing the GEE platform. Achieving high classification accuracy over large region is a complex task and it requires highly stable and robust computational resources for image processing and running the machine learning algorithms. GEE not only made the task easier by handling the requirement for high computational performance but also by providing ready to use advance machine learning algorithms which have their own challenges in other platforms while utilized for satellite image analysis. One of the current limitations of GEE is for preparing ready-to-use map composition and handling vector analysis as performed by external commercial platforms such as ArcGIS.

High-resolution monsoon cropland maps are very important to provide accurate monsoon crop location information for assessing the crop condition and possible policy intervention in case of crop failure. The overall accuracy of 93% achieved in this study, for the binary cropland/non-cropland map, suggests that the combined approach introduced in this research is reliable for monsoon cropland mapping and outperforms that of using only Sentinel-1 radar images, especially in regions dominated by rainfed-dryland crops. The combined approach provides classification accuracy of 90% or more in different agro-ecological regions dominated by diverse crops. There is overall 3% increase in classification accuracy from radar only to radar and optical combined for RF whereas the same for SVM increases from 80% to 90 % for monsoon cropland mapping. The performance of both RF and SVM increases across the AER’s. For both RF and SVM, the classification accuracy improved the most for AER-2 and AER-4&5 whereas it has increased the least for AER-2. Overall, the performance of RF was better than SVM. The performance of SVM decreased mainly in the transition regions from cropland to non-cropland and vice versa.

The ROM proposed here has overcome the challenge of differentiating natural vegetation from monsoon cropland mapped during the peak growth stages in monsoon season. Thus, it has applications for segregating cropland from vegetation cover, and may assist in generating a non-crop mask in regions affected by cloud cover. This study can provide important information for decision makers and researchers as monitoring these crops is a challenging task due to small farm/cropland size and frequent cloud cover during the crop-growing season.

The primary objectives of this study are to fill the gap in monsoon cropland monitoring by:

(1) evaluating Sentinel-1 (S1) radar and a combination of Sentinel-1 (S1) radar and Sentinel-2 (S2) optical data in terms of providing greater accuracy for monsoon cropland mapping.

(2) developing a high resolution, all weather applicable non-crop mask for segregating monsoon cropland from forested and agro-forested (plantation) lands with similar signatures.

The study shows that there is overall 3% increase in classification accuracy from radar (S1) to radar+optical (S1+S2) for RF whereas the same for SVM increases by 10% from 80% to 90% for monsoon cropland mapping. The performance of both RF and SVM increases across the AER's. For both RF and SVM, the classification accuracy improved the most for AER-2 and AER-4&5 whereas it has increased the least for AER-2 using both the classifiers. Overall, the performance of RF was better than SVM. The performance of SVM decreased specifically in the transition regions from cropland to non-cropland and vice versa.

6.1 Future Research

This work has shown promising results for smallholder cropland mapping in major rainfed region in India. However, the spatial and temporal transferability of the developed methodology needs to be tested. While this method was developed for monsoon cropland mapping, it is likely that this method can be used to map smallholder cropland affected by cloud cover in Africa and other regions of the world. Future research will include evaluating the performance of this method for other monsoon years with different weather conditions.

REFERENCES

Akar, Ö.; Güngör, O. Classification of multispectral images using Random Forest algorithm. *J. Geodesy Geoinf.* 2012, *1*, 105–112.

Attri S.;& Tyagi, A; Climate profile of India, Contribution to the Indian Network of Climate Change Assessment (National Communication-II), Ministry of environment and forests, 1501, Environment Monitoring and Research Centre, India Meteorological Department, New Delhi,2010, pp. 1-129.

Beck, H.E.; Zimmermann, N.E.; McVicar, T.R.; Vergopolan, N.; Berg, A.; Wood, E.F. Present and future köppen-geiger climate classification maps at 1-km resolution. *Sci. Data* 2018, *5*, 180214.

Becker-Reshef, I.; Justice, C.; Sullivan, M.; Vermote, E.; Tucker, C.; Anyamba, A.; Small, J.; Pak, E.; Masuoka, E.; Schmaltz, J.; et al. Monitoring Global Croplands with Coarse Resolution Earth Observations: The Global Agriculture Monitoring (GLAM) Project. *Remote Sens.* 2010, *2*, 1589–1609.

Benedetti, R.; Rossini, P. On the use of NDVI profiles as a tool for agricultural statistics: The case study of wheat yield estimate and forecast in Emilia Romagna. *Remote Sens. Environ.* 1993, *45*, 311–326.

Breiman, L. *Machine Learning*; Statistics Department, University of Berkeley.: Berkeley, CA, USA, 2001; Volume 45, pp. 5–32.

Bruinsma,2009 J. Expert Meeting on How to Feed the World in 2050 The Resource Outlook TO 2050: 1 By how much do land, water and crop yields need to increase by 2050?;

Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* 1991, *37*, 35–46.

Carrasco, L.; O'Neil, A.W.; Daniel Morton, R.; Rowland, C.S. Evaluating combinations of temporally aggregated Sentinel-1, Sentinel-2 and Landsat 8 for land cover mapping with Google Earth Engine. *Remote Sens.* 2019, *11*, 288.

Chakraborty, M.; Panigrahy, S.; Rajawat, A.S.; Kumar, R.; Murthy, T.V.R.; Haldar, D.; Chakraborty, A.; Kumar, T.; Rode, S.; Kumar, H.; et al. Initial results using RISAT-1 C-band SAR data. *Curr. Sci.* 2014, *104*, 490–501.

Chakraborty, M.; Panigrahy, S.; Sharma, S.A. Discrimination of rice crop grown under different cultural practices using temporal ERS-1 synthetic aperture radar data. *ISPRS J. Photogramm. Remote Sens.* 1997, *52*, 183–191.

Clevers, J.G.P.W.; Van Leeuwen, H.J.C. Combined use of optical and microwave remote sensing data for crop growth monitoring. *Remote Sens. Environ.* 1996, *56*, 42–51.

Dadhwal, V.K.; Singh, R.P.; Dutta, S.; Parihar, J.S. Remote sensing based crop inventory: A review of Indian experience. *Trop. Ecol.* 2002, *43*, 107–122.

Debats, S.R.; Luo, D.; Estes, L.D.; Fuchs, T.J.; Caylor, K.K. A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. *Remote Sens. Environ.* 2016, *179*, 210–221.

De Alban, J.D.T.; Connette, G.M.; Oswald, P.; Webb, E.L. Combined Landsat and L-band SAR data improves land cover classification and change detection in dynamic tropical landscapes. *Remote Sens.* 2018, *10*, 306.

Defourny, P.; Bontemps, S.; Bellemans, N.; Cara, C.; Dedieu, G.; Guzzonato, E.; Hagolle, O.; Inglada, J.; Nicola, L.; Rabaute, T.; et al. Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sens. Environ.* 2019, *221*, 551–568.

Delrue, J.; Bydekerke, L.; Eerens, H.; Gilliams, S.; Piccard, I.; Swinnen, E. Crop mapping in countries with small-scale farming: A case study for West Shewa, Ethiopia. *Int. J. Remote Sens.* 2013, *34*, 2566–2582.

FAO 2014, Crop Monitoring for Improved Food Security | Asian Development Bank Available online: https://www.adb.org/publications/crop-monitoring-improved-food-security (accessed on Apr 14, 2020).

FAO 2015-2, The economic lives of smallholder farmers |Available online: http://www.fao.org/family-farming/detail/en/c/385065/ (accessed on Apr 14, 2020).

FAO. 2017. The future of food and agriculture – Trends and challenges. Rome. Available online: https://reliefweb.int/report/world/future-food-and-agriculture-trends-and-challenges (accessed on May 6, 2020).

Feng, Q.; Liu, J.; Gong, J. UAV remote sensing for urban vegetation mapping using random forest and texture analysis. *Remote Sens.* 2015, 7, 1074–1094.

Fritz, S.; See, L.; Bayas, J.C.L.; Waldner, F.; Jacques, D.; Becker-Reshef, I.; Whitcraft, A.; Baruth, B.; Bonifacio, R.; Crutchfield, J.; et al. A comparison of global agricultural monitoring systems and current gaps. *Agric. Syst.* 2019, *168*, 258–272.

Gajbhiye, K.S.; Mandal, C. *Agro-Ecological Zones, Their Soil Resource and Cropping Systems*; National Bureau of Soil Survey and Land Use Planning: Nagpur, India, 2000. Available online: http://www.indiawaterportal.org/sites/indiawaterportal.org/files/01jan00sfm1.pdf (accessed on 11 July 2019).

Gatti, A.; Bertolini, A. Sentinel-2 Products Specification Document. 2015. Available online:https://sentinel.esa.int/documents/247904/349490/S2_MSI_Product_Specification.pdf (accessed on 30 June 2019).

Gallup, C.D.; Riker-Coleman, K.E. Earth's climate: Past and future. *Eos, Trans. Am. Geophys. Union* 2001, *82*, 576–576.

Galvin, J.F.P. The weather and climate of the tropics: Part 6 – Monsoons. *Weather* 2008, *63*, 129–137.

Google Developers Image Collection Reductions. Available online: https://developers.google.com/earth-engine/reducers_image_collection (accessed on 19 July 2019).

Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 2017, *202*, 18–27.

Granados-Ramírez, R.; Reyna-Trujillo, T.; Gómez-Rodríguez, G.; Soria-Ruiz, J. Analysis of NOAA-AVHRR-NDVI images for crops monitoring. *Int. J. Remote Sens.* 2004, *25*, 1615–1627.

GRAIN | Hungry for land: small farmers feed the world with less than a quarter of all farmland Available online: https://www.grain.org/article/entries/4929-hungry-for-land-small-farmers-feed-the-world-with-less-than-a-quarter-of-all-farmland (accessed on May 6, 2020).

Haldar, D.; Patnaik, C.; Chakraborty, M. Jute Crop Discrimination and Biophysical Parameter Monitoring Using Multi-Parametric SAR Data in West Bengal, India. *Open Access Lib. J.* 2014, *1*, 1.

Halpert, M.S.; Bell, G.D. Climate Assessment for 1996. *Bull. Am. Meteorol. Soc.* 1997, *78*, S1–S50.

Handique BK, Goswami J, Qadir A, Gupta C, Raju PLN (2016) Rapid assessment of boro paddy infestation by brown plant hopper in Morigaon district, Assam, India using unmanned aerial vehicle. Curr Sci 111:1604–1606

Handique, B.K.; Khan, A.Qadir.; Goswami, C.; Prashnani, M.; Gupta, C.; Raju, P.L.N. Crop Discrimination Using Multispectral Sensor Onboard Unmanned Aerial Vehicle. *Proc. Natl. Acad. Sci. India Sect. A - Phys. Sci.* 2017, *87*, 713–719.

Huang, C.; Davis, L.S.; Townshend, J.R.G. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* 2002, *23*, 725–749.

Inoue, Y.; Kurosu, T.; Maeno, H.; Uratsuka, S.; Kozu, T.; Dabrowska-Zielinska, K.; Qi, J. Season-long daily measurements of multifrequency (Ka, Ku, X, C, and L) and full-polarization backscatter signatures over paddy rice field and their relationship with biological variables. *Remote Sens. Environ.* 2002, *81*, 194–204.

Inglada, J.; Arias, M.; Tardy, B.; Hagolle, O.; Valero, S.; Morin, D.; Dedieu, G.; Sepulcre, G.; Bontemps, S.; Defourny, P.; et al. Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery. *Remote Sens.* 2015, *7*, 12356–12379.

India: Issues and Priorities for Agriculture Available online: https://www.worldbank.org/en/news/feature/2012/05/17/india-agriculture-issues-priorities (accessed on Apr 14, 2020).

Jain, M.; Mondal, P.; DeFries, R.S.; Small, C.; Galford, G.L. Mapping cropping intensity of smallholder farms: A comparison of methods using multiple sensors. *Remote Sens. Environ.* 2013, *134*, 210–223.

Jin, H.; Xu, W.; Li, A.; Xie, X.; Zhang, Z.; Xia, H. Spatially and Temporally Continuous Leaf Area Index Mapping for Crops through Assimilation of Multi-resolution Satellite Data. *Remote Sens.* 2019, *11*, 2517.

Jia, K.; Li, Q.; Tian, Y.; Wu, B.; Zhang, F.; Meng, J. Crop classification using multi-configuration SAR data in the North China plain. *Int. J. Remote Sens.* 2012, *33*, 170–183.

Jiang, D.; Wang, N.B.; Yang, X.H.; Wang, J.H. Study on the interaction between NDVI profile and the growing status of crops. *Chin. Geogr. Sci.* 2003, *13*, 62–65.

Kamusoko, C.; Gamba, J.; Murakami, H. Mapping Woodland Cover in the Miombo Ecosystem: A Comparison of Machine Learning Classifiers. *Land* 2014, *3*, 524–540.

Khullar, D. R; India: A comprehensive geography. 2008, New Delhi: Kalyani Publishers.

Land Use Statistics Information System. Available online: https://aps.dac.gov.in/APY/Index.htm (accessed on 30 June 2019).

Lange, H.; Sippel, S. Machine Learning Applications in Hydrology. In; Springer, Cham, 2020; pp. 233–257.

Larrañaga, A.; Álvarez-Mozos, J. On the Added Value of Quad-Pol Data in a Multi-Temporal Crop Classification Framework Based on RADARSAT-2 Imagery. *Remote Sens*. 2016, *8*, 335.

Liu, J.; Li, W.; Tian, Y. *Automatic Thresholding of Gray-Level Pictures Using Two-Dimension Otsu Method*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, New Jersey, United States, 2002; pp. 325–327.

Liu, D.; Xia, F. Assessing object-based classification: Advantages and limitations. *Remote Sens. Lett.* 2010, *1*, 187–194.

Liaw Merck, A.; Liaw, A.; Wiener, M. *Classification and Regression by RandomForest*. R News, 2002. 2, 18-22. Available online at: http://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf.

Lowder, S.K.; Skoet, J.; Raney, T. The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide. *World Dev.* 2016, *87*, 16–29.

Lone, J.M.; Sivasankar, T.; Sarma, K.K.; Qadir, A.; Raju, P.L.N. Influence of Slope Aspect on above Ground Biomass Estimation using ALOS-2 Data. Int. J. Sci. Res. 2017, 6, 1422–1428.

Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* 2007, *28*, 823–870.

Khamala, E.; Food and Agriculture Organization of the United Nations *Review of the available remote sensing tools, products, methodologies and data to improve crop production forecasts*;2017 ISBN 9789251098400.

Kumari, M.; Murthy, C.S.; Pandey, V.; Bairagi, G.D. Soybean Cropland mapping using Multi-Temporal Sentinel-1 data. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* 2019, *XLII-3/W6*, 109–114.

Kussul, N.; Mykola, L.; Shelestov, A.; Skakun, S. Crop inventory at regional scale in Ukraine: Developing in season and end of season crop maps with multi-temporal optical and SAR satellite imagery. *Eur. J. Remote Sens.* 2018, *51*, 627–636.

Madhusudhan, L. Agriculture Role on Indian Economy. *Bus. Econ. J.* 2015, *6*, 1.

Macedo-Cruz, A.; Pajares, G.; Santos, M.; Villegas-Romero, I. Digital image sensor-based assessment of the status of oat (Avena sativa L.) crops after frost damage. *Sensors* 2011, *11*, 6015–6036.

MacDonald, R.; Hall, F.; Erb, R. The Use of LANDSAT Data in a Large Area Crop Inventory Experiment (LACIE). *LARS Symp.* 1975.

McNairn, H.; Champagne, C.; Shang, J.; Holmstrom, D.; Reichert, G. Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories. *ISPRS J. Photogramm. Remote Sens.* 2009, *64*, 434–449.

Mansaray, L.; Huang, W.; Zhang, D.; Huang, J.; Li, J. Mapping Rice Fields in Urban Shanghai, Southeast China, Using Sentinel-1A and Landsat 8 Datasets. *Remote Sens.* 2017, *9*, 257.

Mapping cropland in smallholder farmer systems in South-Africa using Sentinel-2 imagery | Mémoire UCL Available online: https://dial.uclouvain.be/memoire/ucl/fr/object/thesis%3A13397 (accessed on Apr 14, 2020).

Massey, R.; Sankey, T.T.; Yadav, K.; Congalton, R.G.; Tilton, J.C. Integrating cloud-based workflows in continental-scale cropland extent classification. *Remote Sens. Environ.* 2018, *219*, 162–179.

Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* 2019, *152*, 166–177.

Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of machine-learning classification in remote sensing: An applied review. *Int. J. Remote Sens.* 2018, *39*, 2784–2817.

Mercier, A.; Betbeder, J.; Rumiano, F.; Baudry, J.; Gond, V.; Blanc, L.; Bourgoin, C.; Cornu, G.; Ciudad, C.; Marchamalo, M.; et al. Evaluation of Sentinel-1 and 2 Time Series for Land Cover Classification of Forest–Agriculture Mosaics in Temperate and Tropical Landscapes. *Remote Sens.* 2019, *11*, 979.

Memarian, H.; Balasundram, S.K.; Khosla, R. Comparison between pixel- and object-based image classification of a tropical landscape using Système Pour l'Observation de la Terre-5 imagery. *J. Appl. Remote Sens.* 2013, *7*, 073512.

Mondal, P.; McDermid, S.S.; Qadir, A. A reporting framework for Sustainable Development Goal 15: Multi-scale monitoring of forest degradation using MODIS, Landsat and Sentinel data. *Remote Sens. Environ.* 2020, *237*, 111592.

Mondal, P.; Liu, X.; Fatoyinbo, T.E.; Lagomasino, D. Evaluating Combinations of Sentinel-2 Data and Machine-Learning Algorithms for Mangrove Mapping in West Africa. *Remote Sens.* 2019, *11*, 2928.

Moorthi, S.M.; Oza, M.; Misra, I.; Gambhir, R.K.; Darji, N.P.; Sharma, S.; Jain, D.K.; Dhar, D.; Padia, K.; Ramakrishnan, R.; et al. FASALSoft - An ISRO software framework for crop production forecast using remote sensing data analysis 2014.

Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens*. 2011, 66, 247–259.

Mtibaa, S.; Irie, M. Land cover mapping in cropland dominated area using information on vegetation phenology and multi-seasonal Landsat 8 images. *Euro-Mediterranean J. Environ. Integr.* 2016, *1*.

NFSM: National Food Security Mission. Available online: https://www.nfsm.gov.in/ (accessed on 31 October 2019).

(No Title) Available online: http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf (accessed on Apr 14, 2020).

(No Title) Available online: http://www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_Global_Agriculture.pdf (accessed on Apr 14, 2020).

Okwuashi, O.; Isong, M.; Eyo, E.; Eyoh, A.; Nwanekezie, O.; Olayinka, D.N.; Udoudo, D.O.; Ofem, B. GIS Cellular Automata Using Artificial Neural Network for Land Use Change Simulation of Lagos, Nigeria. *J. Geogr. Geol.* 2012, *4*, p94.

Parihar, J. FASAL concept in meeting the requirements of assessment and forecasting crop production affected by extreme weather events. *Mausam.2016,* 67. 93-104.

Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens*. 2005, *26*, 217–222.

Parthasarathy, B.; Munot, A.A.; Kothawale, D.R. All-India monthly and seasonal rainfall series: 1871–1993. *Theor. Appl. Climatol*. 1994, 49, 217–224.

Pericak, A.A.; Thomas, C.J.; Kroodsma, D.A.; Wasson, M.F.; Ross, M.R.V.; Clinton, N.E.; Campagna, D.J.; Franklin, Y.; Bernhardt, E.S.; Amos, J.F. Mapping the yearly extent of surface coal mining in central appalachia using landsat and google earth engine. *PLoS ONE* 2018, *13*, e0197758.

Pérez-Hoyos, A.; Rembold, F.; Kerdiles, H.; Gallego, J. Comparison of global land cover datasets for cropland monitoring. *Remote Sens.* 2017, *9*.

Pittman, K.; Hansen, M.C.; Becker-Reshef, I.; Potapov, P.V.; Justice, C.O. Estimating global cropland extent with multi-year MODIS data. *Remote Sens.* 2010, *2*, 1844–1863.

Powers, D.M. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. *J. Mach. Learn. Technol.* 2011, *1*, 37–63

Prasanna, V. Impact of monsoon rainfall on the total foodgrain yield over India. *J. Earth Syst. Sci.* 2014, *123*, 1129–1145.

Qadir, A.; Mondal, P. Synergistic Use of Radar and Optical Satellite Data for Improved Monsoon Cropland Mapping in India. *Remote Sens.* 2020, *12*, 522.

Quynh Trang, N.T.; Toan, L.Q.; Huyen Ai, T.T.; Vu Giang, N.; Viet Hoa, P. Object-Based vs. Pixel-Based Classification of Mangrove Forest Mapping in Vien An Dong Commune, Ngoc Hien District, Ca Mau Province Using VNREDSat-1 Images. *Adv. Remote Sens.* 2016, *05*, 284–295.

Qian, Y.; Zhou, W.; Yan, J.; Li, W.; Han, L. Comparing Machine Learning Classifiers for Object-Based Land Cover Classification Using Very High Resolution Imagery. *Remote Sens.* 2014, *7*, 153–168.

Ranjan, A.K.; Parida, B.R. Paddy acreage mapping and yield prediction using sentinel-based optical and SAR data in Sahibganj district, Jharkhand (India). *Spat. Inf. Res.* 2019, *27*, 399–410.

Rakwatin P, Prakobya A, Sritarapipat T, Khobkhun B, Pannangpetch K, Sobue S, Oyoshi K, Okumura T, T.N. Rice Crop Monitoring In Thailand Using Field Server And Satellite Remote Sensing. In Proceedings of the Proceedings of the Expert Meeting on Crop Monitoring for Food Security; 2014; pp. 79–84.

Ricciardi, V.; Ramankutty, N.; Mehrabi, Z.; Jarvis, L.; Chookolingo, B. How much of the world's food do smallholders produce? *Glob. Food Sec.* 2018, *17*, 64–72.

Rhee, J.; Im, J. Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. Agric. For. Meteorol. 2017, 237–238, 105–122.

Rosenzweig, C.; Iglesias, A.; Yang, X.B.; Epstein, P.; Chivian, E. Climate Change and Extreme Weather Events; Implications for Food Production, Plant Diseases, and Pests. *Glob. Chang. Hum. Heal.* 2001, *2*, 90–104.

Roy, P.S.; Sharma, K.P.; Jain, A. Stratification of density in dry deciduous forest using satellite remote sensing digital data—An approach based on spectral indices. *J. Biosci.* 1996, *21*, 723–734.

Rodriguez-Galiano, V.F.; Chica-Olmo, M.; Abarca-Hernandez, F.; Atkinson, P.M.; Jeganathan, C. Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sens. Environ.* 2012, *121*, 93–107.

Roy, A.; Inamdar, A.B. Multi-temporal Land Use Land Cover (LULC) change analysis of a dry semi-arid river basin in western India following a robust multi-sensor satellite image calibration strategy. *Heliyon* 2019, *5*.

Rußwurm, M.; Korner, M. Temporal Vegetation Modelling Using Long Short-Term Memory Networks for Crop Identification from Medium-Resolution Multi-spectral Satellite Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, New Jersey, United States, 2017; pp. 1496–1504.

Sentinel-2 User Handbook - Document Library | Sentinel Online Content Available online: https://sentinel.esa.int/web/sentinel/document-library/content/-/article/sentinel-2-user-handbook (accessed on May 6, 2020).

Scanlon, T.M.; Albertson, J.D.; Caylor, K.K.; Williams, C.A. Determining land surface fractional cover from NDVI and rainfall time series for a savanna ecosystem. *Remote Sens. Environ.* 2002, *82*, 376–388.

Shang, J.; McNairn, H.; Champagne, C.; Jiao, X. Contribution of multi-frequency, multi-sensor, and multi-temporal radar data to operational annual crop mapping. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS); 2008; Vol. 3.

Shang, J.; McNairn, H.; Champagne, C.; Jiao, X. Application of Multi-Frequency Synthetic Aperture Radar (SAR) in Crop Classification. In *Advances in Geoscience and Remote Sensing*; InTech, 2009.

Shang, R.; Liu, R.; Xu, M.; Liu, Y.; Dash, J.; Ge, Q. Determining the start of the growing season from MODIS data in the Indian Monsoon Region: Identifying available data in the rainy season and modeling the varied vegetation growth trajectories. *Remote Sens.* 2018, *11*, 939.

Singha, M.; Dong, J.; Zhang, G.; Xiao, X. High resolution paddy rice maps in cloud-prone Bangladesh and Northeast India using Sentinel-1 data. *Sci. Data* 2019, *6*, 26.

Singha, M.; Wu, B.; Zhang, M. An object-based paddy rice classification using multi-spectral data and crop phenology in Assam, Northeast India. *Remote Sens.* 2016, *8*, 479.

Sirro, L.; Häme, T.; Rauste, Y.; Kilpi, J.; Hämäläinen, J.; Gunia, K.; de Jong, B.; Pellat, F.P. Potential of different optical and SAR data in forest and land cover classification to support REDD+ MRV. *Remote Sens.* 2018, *10*, 942.

Sivasankar, T.; Kumar, D.; Srivastava, H.S.; Patel, P. Advances in radar remote sensing of agricultural crops: A review. *Int. J. Adv. Sci. Eng. Inf. Technol.* 2018, *8*, 1126–1137.

Sivasankar, T.; Lone, J.M.; Sarma, K.K.; Qadir, A.; Raju, P.L. The potential of multi-frequency multipolarized ALOS-2/PALSAR-2 and Sentinel-1 SAR data for aboveground forest biomass estimation. *Int. J. Eng. Technol.* 2018, *10*, 797–802

Sivasankar, T.; Lone, J.M.; K., S.K.; Qadir, A.; N., R.P.L. Estimation of Above Ground Biomass Using Support Vector Machines and ALOS/PALSAR data. *Vietnam J. Earth Sci.* 2019, *41*, 95–104.

Skriver, H.; Mattia, F.; Satalino, G.; Balenzano, A.; Pauwels, V.R.N.; Verhoest, N.E.C.; Davidson, M. Crop Classification Using Short-Revisit Multitemporal SAR Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2011, *4*, 423–431.

Skakun, S.; Kussul, N.; Shelestov, A.Y.; Lavreniuk, M.; Kussul, O. 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.* 2016, *9*, 3712–3719.

Sun, C.; Bian, Y.; Zhou, T.; Pan, J. Using of multi-source and multi-temporal remote sensing data improves crop-type mapping in the subtropical agriculture region. *Sensors (Switzerland)* 2019, *19*.

Salah, M. A survey of modern classification techniques in remote sensing for improved image classification. *J. Geomat.* 2017, *11*, 1–21

Tavares, P.A.; Beltrão, N.E.S.; Guimarães, U.S.; Teodoro, A.C. Integration of sentinel-1 and sentinel-2 for classification and LULC mapping in the urban area of Belém, eastern Brazilian Amazon. *Sensors* 2019, *19*, 1140.

Tian, F.; Wu, B.; Zeng, H.; Zhang, X.; Xu, J. Efficient Identification of Corn Cultivation Area with Multitemporal Synthetic Aperture Radar and Optical Images in the Google Earth Engine Cloud Platform. *Remote Sens.* 2019, *11*, 629.

Timmermans, Alice. Mapping cropland in smallholder farmer systems in South-Africa using Sentinel-2 imagery. Faculté des bioingénieurs, Université catholique de Louvain, 2018.

Toosi, N.B.; Soffianian, A.R.; Fakheran, S.; Pourmanafi, S.; Ginzler, C.; Waser, L.T. Comparing different classification algorithms for monitoring mangrove cover changes in southern Iran. *Glob. Ecol. Conserv.* 2019, *19*, e00662.

Topaloğlu, R.H.; Sertel, E.; Musaoğlu, N. Assessment of classification accuracies of Sentinel-2 and Landsat-8 data for land cover/use mapping. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives; International Society for Photogrammetry and Remote Sensing, 2016; Vol. 41, pp. 1055–1059.

Thanh Noi, P.; Kappas, M. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors* 2017, *18*, 18.

Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 1979, *8*, 127–150.

Tutorials-Sentinel-1 Toolbox-Sentinel Online. Available online: https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-1/tutorials (accessed on 30 June 2019).

Van der Mensbrugghe, D.; Osorio Rodarte, I.; Burns, A.; Baffes, J. How to feed the world in 2050: Macroeconomic environment, commodity markets - A longer term outlook. 2009.

Van Tricht, K.; Gobin, A.; Gilliams, S.; Piccard, I. Synergistic Use of Radar Sentinel-1 and Optical Sentinel-2 Imagery for Crop Mapping: A Case Study for Belgium. *Remote Sens.* 2018, *10*, 1642.

Verma, A.; Kumar, A.; Lal, K. Kharif crop characterization using combination of SAR and MSI Optical Sentinel Satellite datasets. *J. Earth Syst. Sci.* 2019, *128*, 230.

Waldner, F.; De Abelleyra, D.; Verón, S.R.; Zhang, M.; Wu, B.; Plotnikov, D.; Bartalev, S.; Lavreniuk, M.; Skakun, S.; Kussul, N.; et al. Towards a set of agrosystem-specific cropland mapping methods to address the global cropland diversity. *Int. J. Remote Sens.* 2016, *37*, 3196–3231.

Wang, D.; Su, Y.; Zhou, Q.; Chen, Z. Advances in research on crop identification using SAR. In Proceedings of the 2015 4th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Istanbul, Turkey, 20–24 July 2015; Institute of Electrical and Electronics Engineers Inc.: Piscataway, New Jersey, United States, 2015; pp. 312–317.

World Bank-2012, India: Issues and Priorities for Agriculture Available online: https://www.worldbank.org/en/news/feature/2012/05/17/india-agriculture-issues-priorities (accessed on May 6, 2020).

Woodhouse, I.H. *Introduction to Microwave Remote Sensing*; Taylor & Francis: Didcot, UK, 2006; ISBN 9780415271233

Whitcraft, A.K.; Vermote, E.F.; Becker-Reshef, I.; Justice, C.O. Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations. *Remote Sens. Environ.* 2015, *156*, 438–447.

Wulder, M.A.; Loveland, T.R.; Roy, D.P.; Crawford, C.J.; Masek, J.G.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Belward, A.S.; Cohen, W.B.; et al. Current status of Landsat program, science, and applications. *Remote Sens. Environ.* 2019, *225*, 127–147.

Wulder, M.A.; White, J.C.; Goward, S.N.; Masek, J.G.; Irons, J.R.; Herold, M.; Cohen, W.B.; Loveland, T.R.; Woodcock, C.E. Landsat continuity: Issues and opportunities for land cover monitoring. *Remote Sens. Environ.* 2008, *112*, 955–969.

Zafari, A.; Zurita-Milla, R.; Izquierdo-Verdiguier, E. Integrating support vector machines and random forests to classify crops in time series of Worldview-2 images. In Proceedings of the Image and Signal Processing for Remote Sensing XXIII; Bruzzone, L., Bovolo, F., Benediktsson, J.A., Eds.; SPIE, 2017; Vol. 10427, p. 34.

Appendix A

PERMISSION

Some parts of this Some parts of this thesis was published in an open access MDPI journal. As per MDPI authorship rights, as author of this work, I retain the right to include this manuscript within this thesis.More can be found at https://www.mdpi.com/authors/rights