Land cover mapping of Sentinel-1 time-series data using deep learning techniques

Author One[1*†], Author Two[2†], Author Three[2], and Author Four[1,2]

[1]_

[2]_

[*]Corresponding author. Email: email@email.com

[†]These authors contributed equally to this work.

Abstract

Spatial and non-spatial data plays important role in social and economic development of a country. Land use and land cover (LULC) maps are one such data type used for retrieving land information and detecting changes in the environment. Several recent research have used deep learning approaches to generate LULC maps from various remote sensing image formats, and have correctly recognized several land cover elements such as built-up areas, forest, water, and so on. Deep learning approaches, on the other hand, need to be tested on spaceborne radar image. The current research will close the gap in classifying Sentinel-1 synthetic aperture radar (SAR) images using four state-of-the-art deep learning models that have largely been employed in prior research for classifying other satellite image formats. Training samples of five primary classes for ground-truth data were obtained by using Google Earth and Sentinel-2 images as a reference map. For segmenting Sentinel-1 images, four deep learning semantic segmentation models were chosen and fine-tuned: U-Net, SegNet, FCN, and DeepLabV3+. All of the models did exceptionally well, with total accuracy ranging from 80% to 85%. U-Net and FCN fared well in classifying all five land cover features, with mean inter-section over union (IoU) of 0.67 and 0.46, respectively. The results reveal that the Sentinel-1 dual-pol images can be utilised to categorise the primary land cover features using a self-generated reference map.

Keywords: Sentinel-1, deep learning, semantic segmentation, ground-truth image, land use land cover, synthetic aperture radar

1 Introduction

In terms of spaceborne/airborne images, classification entails assigning classes/labels to each spatial unit or pixels. The classification of human actions (such as buildings, bridges/highways, rail tracks, airports, dams and reservoirs) and natural components (such as mountains, plateaus, hills, minerals, and forests) on the landscape over time using established scientific and statistical methods of analysis of significant materials is known as land cover mapping or land use land cover (LULC) [1]. In simple

words, land cover refers to the ground's surface cover, which includes things like infrastructural development, plants, water, fallow land, and so on. The cornerstone for tasks like thematic mapping and change detection analysis is land cover identification. Natural resource management [2], wildlife habitat protection [3], urban expansion/encroachment [4], natural catastrophe (tornadoes, flooding, volcanic, earthquake, fire) [5], and target detection [6] are all covered by LULC maps. LULC maps helps in policy making and development planning at local, regional and national levels [7]. With LULC map one can create national level seamless data for information system development, LULC monitoring system, urban planning assessment, quantification of storm water run off, estimating irrigation water requirements [8]. Although remote sensing can be utilised to give timely and precise land cover information, optical imagery is still commonly used in land cover mapping. Land cover mapping has been frequently used on multispectral satellite data for the past decade, although most satellite images are impacted by climate and weather conditions, making time-series analysis difficult.

Due to the availability of user-friendly programming tools, high-end consumer computing power, and high-resolution optical/synthetic aperture radar (SAR) satellite image data, data science and remote sensing techniques have become increasingly popular in recent years. Change detection in land cover can also be determined by using historical satellite image dataset [9]. Although LULC maps created from satellite images have biases due to image quality and resolution, incorrectly annotated groundtruth datasets, variability in the study area, minimum mapping unit, and other factors [10], these flaws can limit and complicate how digital LULC maps should be used for applications. However, accuracy indicators can provide information on the quality of a map data set as well as how to use these products efficiently.

1.1 Capabilities and applications of the Sentinel-1 SAR Image

Radar remote sensing from space became a reality with the launch of Seasat in 1972. Since its launch in 2014, the Sentinel 1 mission has piqued the interest of numerous scholars and professionals, and many segmentation algorithms for Sentinel-1 images have been published[11] [12] [13]. Sentinel-1 is a constellation of two polar-orbiting satellites that perform C-band synthetic aperture radar imaging day and night (central frequency: 5.405 GHz) [14], enabling them to acquire imagery regardless of the weather. The European Space Agency's (ESA) Sentinel-1 Earth observation satellite has even greater capabilities, with a five-day revisit period and a 10 m spatial resolution, and unlike other SAR data (e.g., TerraSAR-X, ALOS2, RADARSAT-2, COSMO-Sky Med), it is publicly available under an open licence. Stripmap (SM), Interferometric Wide Swath (IW), Extra Wide Swath (EW), and Wave (WV) are the four imaging modes available, each with differing processing levels, spatial resolutions (down to 5 m), and coverage (up to 400 km). Sentinel-1A (launched on 3 April 2014) and Sentinel-1B (launched on 25 April 2016) are interferometric wide-swath satellites that can complete a cycle at the Equator in six days[15]. In order to maximise the Signal-to-Noise (S/N) ratio while acquiring any remote sensing images, whether spaceborne or airborne, users must evaluate how much of the recorded signal appears to be pixel useable information and how much is undesired distortion or noise. Similarly, SAR images contain speckle noises that are more complex than noise in other remote

sensing data, because speckle is a granular noise that exists inherently and degrades the quality of SAR images; however, the effect of speckle decreases for high resolution images as the number of elemental scatters within a resolution cell decreases; however, the effect of speckle decreases for high resolution images as the number of elemental scatters within a resolution cell decreases [16]. Despite aforementioned capabilities, the Sentinel-1 image has a wide range of applications, including crop monitoring. [17], marine application like ship detection [18], land monitoring [19], disaster management like landslides [20], estimating soil characteristic [21], estimating forest above ground biomass [22] etc. Computer vision is essential for retrieving the aforementioned applications from Sentinel-1 images[23]. As a result of various successful applications in the computer vision community, deep learning in remote sensing is gaining popularity.

1.2 Deep learning techniques used for segmentation on Sentinel-1 image

Deep learning (DL) [24] has advanced at a breakneck pace in recent years, generating remarkable success in a variety of sectors. DL is important for classifying land use and land cover in geographic remote sensing data because it is based on pixel-by-pixel classification/segmentation and has hierarchical feature learning capabilities, which assigns a label to the type of object present in the data. Many deep neural network algorithms have been developed for classifying optical sensor data, which primarily operate in the electromagnetic (EM) spectrum with wavelengths ranging from 0.4 to 15 micrometres, whereas microwave sensors operate in the EM spectrum with wavelengths ranging from 0.1 cm to 100 cm [25]. DL replicates the working pattern of the human brain; deep learning networks aid in data processing and pattern recognition, and they are capable of learning from unstructured or unlabeled data for decision-making purposes, which is why the name "deep neural network" was used.

Machine learning and deep learning techniques have been widely used for classifying boreal landscapes on multispectral satellite images. However, few research have used these data to evaluate the performance of land cover studies using deep learning algorithms via. SegNet [26], DeepLabV3+ [27], PSPNet [28], BiSeNet [29], FRRN-B [30], U-Net [31], Y-Net [32], Convolutional Neural Network (CNN) [33], Recurrent Neural Network (RNN) [34], Denoising Auto Encoder (DAE) [35], Fully Convolutional Network (FCN) [36] etc., on complex SAR satellite images. SAR images are more complicated than optical images, collecting training samples or making acceptable annotated images for supervised deep learning based techniques from real data is impossible. Based on the aforementioned deep learning models that have been applied on various remote sensing image formats by other researchers, four state-of-the-art semantic segmentation models were chosen for classifying selected land cover features on Sentinel-1 image: UNet, Fully Convolution Network (FCN), SegNet and DeeplabV3+.

1.2.1 U-Net

U-Net was initially proposed by Ronneberger for segmenting biomedical image [31] and there after it was widely used by many researchers for segmenting satellite images for various applications like building detection on SAR image [37], land cover classification on optical and SAR image [38], precise

tree recognition on high resolution optical image [39] and cloud detection [40]. The architecture of U-Net model is made up of 'U' shape and consists of two main paths, contraction path also known as encoder and expansion path also known has decoder[31]. Encoder extracts the image's features information with the help of convolutional and max pooling layers whereas decoder uses transposed convolutional layer to achieve exact localization. It does not have any dense layer thus represents fully convolutional layer.

1.2.2 Fully Convolution Network (FCN)

Long et al. first proposed Fully Convolution Network in 2015. Instead of the dense layer used in traditional CNN, FCN uses a fully connected layer at the end of their network (Convolutional Neural Network). Convolutions blocks of 2D convolutional layer(Conv2D) that are stacked in the network, pooling layers used for dimensionality reduction of the image features dropout, and batch normalisation, which is a type of regularisation that avoids overfitting, are all part of the FCN framework. In order to incorporate non-linearity into the model, an activation function is also incorporated [36]. One of the biggest drawbacks through in downsampling in an FCN is that it reduces the input resolution by a significant portion, making it impossible to replicate finer features during upsampling, even when utilising complex techniques like Transpose Convolution. As a result, we get a rough result. Adding 'skip connections' from earlier layers in the Upsampling step and summing the two feature maps is one technique to cope with issue. Later layers can use this information to establish accurate segmentation borders thanks to the skip connections. Local predictions with almost accurate global (spatial) structure result from this combination of fine and coarse layers.

1.2.3 SegNet

Badrinarayanan et al., proposed SegNet model in 2015 for semantic segmentation. SegNet architures had encoder and decorder part to generate feature map from an image. Encoder part is a VGG-16 [41] without fully connected layers, consists of convolutional layers followed by batchnormalization regulation, max pooling layers and Relu activation function while in decoder part up-samles the input images' feature map. To create dense feature maps, these feature maps are convolved with a trainable decoder filter bank. Each of these maps is then subjected to a batch normalising procedure. Although its encoder input contains three channels, the decoder corresponding to the first encoder (closest to the input image) creates a multi-channel feature map (RGB). Unlike the other decoders in the network, which produce feature maps with the same number of sizes and channels as their encoder inputs, this decoder produces feature maps with the same number of sizes and channels as the encoder inputs. A trainable soft-max classifier is given the high-dimensional feature representation at the output of the final decoder. This soft-max algorithm classifies each pixel separately [42].

1.2.4 DeepLabV3+

In 2017, Cehn et al. proposed the DeepLab architecture [43]. The image is passed through many convolutional blocks in this semantic segmentation model, and instead of a pooling layer, ASPP

(AtrousSpatial Pyramid Pooling) is used, which is the same as transpose convolution or fractionally strided convolution (for stride less than one)[43]. Over the years, DeepLab has gone through several versions, the most recent of which is DeepLabV3+, which adds a few modules to the original DeepLab model. Image level features are also sent on to the ASPP module in this DeepLabV3+ module, which also incorporates a decoder module to refine segmentation results [44].

1.3 Motivation

Availability of ground-truth annotated images for land cover mapping have been an issue especially for Indian regions because of continuous development and it's dense and heterogeneous land cover features[45]. Mehra et. al, [46] used google earth engine and manually annotated four land cover features namely urban, water, forest and agriculture for model evaluation, they have used IOU methods and found Pyramid Scene Parsing Network (PSP) to be the best model for classifying the four classes with mean IoU (0.64) and weighted IoU (0.88), The freely available satellite data has increased the interest of researchers to explore the field. Sentinel-1 C-band SAR system have ability to penetrate radar waves through volumetric objects which provide more and accurate information on land cover changes and it can be operated during day and night. Information on LULC is life-sustaining component in forming policies regarding economic, agriculture, population and environmental issues at national, regional and global scales. This research will narrow this gap as we will be self generating the annotated dataset of various land cover features from sentinel-2 optical image using machine learning techniques (random forest) [47], this generated annotated datatset will then be used in deep learning models for classifying Sentinel-1 SAR image. We have taken five land cover features (Table 4) and applied semantic segmentation models on Sentinel-1 SAR image. Also, the properties of different features on land changes with season so do the backscattering properties of SAR image, considering this time series dataset of Sentinel-1 image were considered (See Table 3) for capturing the variability. This research will primarily concentrate on selecting a limited number of land cover features from Sentinel-1 SAR images in order to generate an appropriate land cover map. This was accomplished by first selecting training samples or land cover classes from optical data and then using random forest machine learning techniques to classify Sentinel-1 images. Most of the information related to use of various deep learning techniques on Sentinel-1 satellite images for land use land cover mapping are discussed in this research.

1.4 Contribution

The following is a summary of our contributions:

- Generated ground-truth labels from Sentinel-2 images by collecting training samples and using random forest machine learning technique. This annotated image were than verified using google earth images and using data from OpenStreetMap.
- Applied four state-of-the-art semantic segmentation models (U-Net, SegNet, FCN and DeepLabV3+) for land cover mapping using Sentinel-1 SAR images, including a wide variety of techniques. We offered recommendations for the most accurate and efficient models i.e., U-Net.

- The ability of deep learning models combined with Sentinel-1 SAR images for accurate land cover mapping in the mixed and heterogeneous boreal land of India was proved in this study. These findings can be used as a starting point for creating new , specialised methods to SAR images.

1.5 Organization of the paper

Figure1 shows the organisation of the paper, which is divided into five sections. The first section will introduce SAR images and deep learning models, as well as provide motivation for current research and contributions. The literature review in Section 2 will provide an overview of publicly available SAR pictures as well as past research by other academics on numerous deep neural techniques for land cover mapping. Section 3 will describe the suggested methodology, which will include pre-processing of Sentinel-1 images, the development of ground-truth labels for land cover features, and the use of several segmentation models. Section 4 will give the results of all of the segmentation models employed in this research. In section 5, we will conclude our findings as well as potential future directions for mapping land cover features using various forms of SAR image data with high image resolution and the necessity of a properly-labeled ground-truth image.

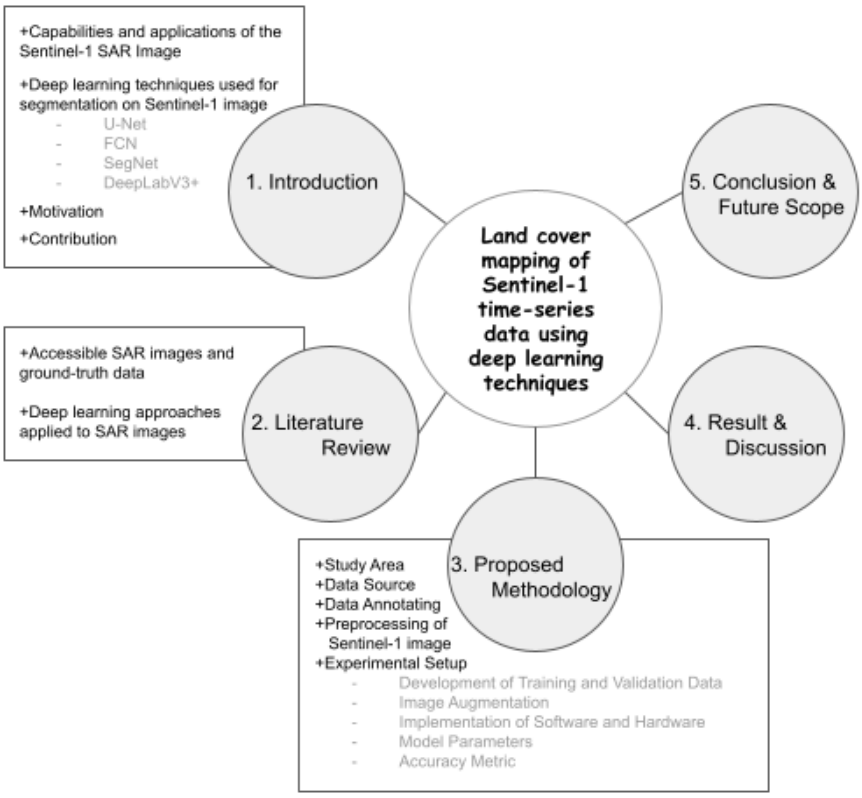

Figure 1: Organization of the paper

2 Literature Review

There are numerous studies that will be covered in this section that use deep learning techniques to deploy various applications on SAR images. This section will be divided into two parts: the first will cover accessible SAR images and ground-truth data, and the second will detail several deep learning approaches applied to SAR images.

2.1 Accessible SAR images and ground-truth data

The freely available satellite data has increased the interest of researchers to explore the remote sensing field. The Sentinel-1 C-band SAR system can penetrate radar signals through volumetric objects, providing better and more precise information on land cover changes, and it can operate at any time of day or night. Table **??** categorizes the approaches used to build datasets for ground-truth images with various land cover features, as well as their availability. Adrian et al. [48] labelled 13 crops including urban, water and soil with the help of field data and WorldView-3 (WV3) images, Solórzano et al. [38] created 10 land cover features by using Sentinel-2 images, Planet images, VHR (Very High Resolution) images from google earth and also collected field data, Šćepanović et al. [49] has used freely available high resolution CORINE map [50] having 5 land cover features which is mainly generated for European and seven cooperating countries [50]; Mehra et al. [46], Garg et al. [51] and Gargiulo et al, [52] used 4 class labels mostly having urban, water, forest/vegetation as common class. Mehra et al. [46] applied CART (Classification And Regression Tree) classifier on LANDSAT-8 images to create annotated images, although barren land, which is plainly detectable in the Indian region, was not included in their research. The authors of the papers [51] [53] [52] and [54] did not specify the source of their ground-truth data and did not go into great detail about land cover aspects. Emek et al. [37] used one class for building detection which is taken from SpaceNet database. Scepanovic et al. [49] used CORINE land cover map which as a annotated dataset for land cover features with 100 m X 100 m spatial resolution for across the European territory and 20m X 20m for Finland region but like green urban area are included in forest. Most land cover features, such as vegetation, soil, and water bodies, change with time and season, and can also be altered by natural and man-made disasters. These small scale changes can easily be detected by generating a large number of annotated images with the help of high resolution optical images (such as WV3). Our research has focused on generating self-annotated images with the help of freely available Sentinel-2 image and applying machine learning techniques to classify these changes (Table 4).

Ref No / Year	Objective	Source Image	Annotated Image	Class Labels	DL Approach	Advantages	Limitation
[46] 2020	Classifying multi-temporal SAR data	Sentinel-1 Single Pol (VV)	LANDSAT-8/ CART clasiifer	4 classes: urban, forest, water, agriculture	Enc-Dec, FPN, U-Net, PSP, DeepLabv3+ PAN	Pyramid Scene Parsing Network (PSP) is the best network overall in terms of both mean IoU (0.64431) and weighted IoU (0.88059)	Bare field class is not included in the study and only VV pol is considered in the network
2021	Semantic segmentation of PolSAR image data using advanced deep learning model	UAVSAR Quad-Pol (2m/px)	No information available	4 classes: urban, ground, water, forest	RF, KNN, SVM, DeepLabV3+	DeepLabv3+ clocked the highest pixel accuracy of 83.51% and 87.78% while among the machine learning algorithms RF classifier achieved the highest overall pixel accuracy of 77.92%	Used ImageNet data for transfer learning which is not related to SAR image
2020	Urban Scene Segmentation from High-Resolution SAR Data	Miranda-35 (0.15m/px)	OSM data for buildings and roads and government -created swisstopo swiss TLM3Dmap	3 classes: building, road, other	U-Net	pixel accuracy of 95.2% and a mean IoU of 74.7% with data collected over a region of merely 2.2 km2	Road class was misclassified

Ref No / Year	Objective	Source Image	Annotated Image	Class Labels	DL Approach	Advantages	Limitation
2021	Sentinel SAR-optical fusion for crop type mapping	Fusion of Setnienel 1 (VV-VH) and Sentinel-2 optical	WorldView-3 image and field data	13 classes: water, foxtail, cocklebur, sorghum, residue, forest, switchgrass, ragweed, sudangrass, soil, urban, corn, and soybean	3D U-Net, 2D U-Net, SegNet	With a score of 0.99, the image fusion strategy for the 3D U-Net has the highest overall accuracy.	The study was used for crop classification having water and build up classes and also involded field data collection
2019	Semantic Segmentation using Deep Learning: A case of study in Albufera Park, Valencia	Sentinel- 1 (VV-VH) and Sentinel-2 optical	No information available	4 classes: water, vegetation, bare soil, others	FPN, Linknet, U-Net	U-Net showed good result with 0.9 overall accuracy and water and the vegetation pixels are clearly recognized with dual polarimetry	Failed to recognize bare soil class
2020	Building detection from SAR images using U-Net	SAR- Capella Space (50cm/px) Optical- Maxar's WorldView 2 satellite (50cm/px)	SpaceNet, 2020	1 class: building	U-Net	Buildings were easily detected d with the help of U-Net Network	Only one class is considered and tested on only very high spatial resolution SAR data

Ref No / Year	Objective	Source Image	Annotated Image	Class Labels	DL Approach	Advantages	Limitation
2021	Wide-Area Land Cover Mapping with Sentinel-1 Imagery using Deep Learning Semantic Segmentation Models	Sentinel -1 Dual Pol (VV and VH)	high-resolution CORINE map	5 classes: urban, forested areas, water bodies, agriculture, petland	BiSeNet, SegNet, Mobile U-Net, DeepLabV3+ FRRN-B, PSPNet, FC-DenseNet	Use of pre-trained CNNs and countrywide classification experiment using Sentinel-1 IW mode SAR data, reaching nearly 93% overall classification accuracy with the best performing models (SegNet and FCDenseNet).	testing multitemporal approaches, data fusion, and employing DEM models, very high-resolution SAR imagery, as well as developing models specifically for SAR
2020	Integration of Sentinel-1and Sentinel-2 Data for Land Cover Mapping Using W-Net	->Sentinel-1 (VH), Sentinel-2 (NDVI and NDWI) ->R: VH, G: NDVI, B: MNDWI	No information available	3 classes: vegetation, bare soil , water	W-Net, SegNet, FPN, LinkNet, U-Net	Proposed W-Net network achieved 0.91 accuracy and 0.88 precision	Less classes considered

Ref No / Year	Objective	Source Image	Annotated Image	Class Labels	DL Approach	Advantages	Limitation
2021	LULC with U-Net using Sentinel 1 and 2 image	Sentinel-1 (VV-VH) and Sentinel-2 (BGR and NIR)	Field data including Sentinel-2 image, planet images and google earth images	10 classes: aquatic vegetation, human settlements, old-growth forest, secondary forest, roads, old-growth plantations, soil, water, young plantations, grasslands /agriculture,	U-Net, RF	the highest accuracy was achieved using the U-net with the MS + SAR bands as inputs, followed very closely by MS U-net and lastly by SAR U-net.	With SAR data, the overall accuracy was 72%

Table 2: Deep learning approaches applied to SAR images

S.No.	Reference	Year	Objective	Source Image	Annotated Image	Class Labels	DL Approach	Advantages	Limitation
1	\cite{mehra2020novel}	2020	Classifying multi-temporal SAR data	Sentinel-1 Single Pol (VV)	LANDSAT-8/ CART clasiifer	4 classes: urban, forest, water, agriculture	Enc-Dec, FPN, U-Net, PSP, DeepLabv3+, PAN	Pyramid Scene Parsing Network (PSP) is the best network overall in terms of both mean IoU (0.64431) and weighted IoU (0.88059)	Bare field class is not included in the study and only VV pol is considered in the network
2	\cite{adrian2021sentinel}	2021	Sentinel SAR-optical fusion for crop type mapping	Fusion of Setmienel 1 (VV-VH) and Sentinel-2 optical	WorldView-3 image and field data	13 classes: water, foxtail, cocklebur, sorghum, residue, forest, switchgrass, ragweed, sudangrass, soil, urban, corn, and soybean	3D U-Net, 2D U-Net, SegNet	With a score of 0.99, the image fusion strategy for the 3D U-Net has the highest overall accuracy.	The study was used for crop classification having water and build up classes and also involded field data collection
3	\cite{vscepanovic2021wide}	2021	Wide-Area Land Cover Mapping with Sentinel-1 Imagery using Deep LearningSemantic Segmentation Models	Sentinel -1 Dual Pol (VV and VH)	high-resolution CORINE map	5 classes: urban, forested areas, water bodies, agriculture, petland	BiSeNet, SegNet, Mobile U-Net, DeepLabV3+, FRRN-B, PSPNet, FC- DenseNet	Use of pre-trained CNNs and countrywide classification experiment using Sentinel-1 IW mode SAR data, reaching nearly 93% overall classification accuracy with the best performing models (SegNet and FCDenseNet).	testing multitemporal approaches, data fusion, and employing DEM models, very high-resolution SAR imagery, as well as developing models specifically for SAR
4	\cite{solorzano2021land}	2021	LULC with U-Net using Sentinel 1 and 2 image	Sentinel-1 (VV-VH) and Sentinel-2 (BGR and NIR)	Field data including Sentinel-2 image, planet images and google earth images	10 classes: aquatic vegetation, human settlements, old-growth forest, secondary forest, roads, old-growth plantations, soil, water young plantations, grasslands/agriculture,	U-Net, RF	the highest accuracy was achieved using the U-net with the MS + SAR bands as inputs, followed very closely by MS U-net and lastly by SAR U-net.	With SAR data, the overall accuracy was 72%
5	\cite{garg2021semantic}	2021	Semantic segmentation of PolSAR image data using advanced deep learning model	UAVSAR Quad-Pol (2m/px)	No information available	4 classes: urban, ground, water, forest	RF, KNN, SVM, DeepLabV3+	DeepLabv3+ clocked the highest pixel accuracy of 83.51% and 87.78% while among the machine learning algorithms RF classifier achieved the highest overall pixel accuracy of 77.92%	Used ImageNet data for transfer learning which is not realated to SAR image
6	\cite{wang2020hr}	2020	Urban Scene Segmentation from High-Resolution SAR Data	Miranda-35 (0.15m/px)	OSM data for buildings and roads and government-created swisstopo swissTLM3D map	3 classes: building, road, other	U-Net	pixel accuracy of 95.2% and a mean IoU of 74.7% with data collected over a region of merely 2.2 km2	Road class was misclassified
7	\cite{gargiulo2019semantic}	2019	Semantic Segmentation using Deep Learning: A case of study in Albufera Park, Valencia	Sentinel- 1 (VV-VH) and Sentinel-2 optical	No information available	4 classes: water, vegetation, bare soil, others	FPN, Linknet, U-Net	U-Net showed good result with 0.9 overall accuracy and water and the vegetation pixels are clearly recognized with dual polarimetry	Failed to recognize bare soil class
8	\cite{gargiulo2020integration}	2020	Integration of Sentinel-1 and Sentinel-2 Data for Land Cover Mapping Using W-Net	->Sentinel-1 (VH), Sentinel-2 (NDVI and NDWI) ->R: VH, G: NDVI, B: MNDWI	No information available	3 classes: vegetation, bare soil , water	W-Net, SegNet, FPN, LinkNet, U-Net	Proposed W-Net network achieved 0.91 accuracy and 0.88 precision	Less classes considered
9	\cite{emek2020building}	2020	Building detection from SAR images using U-Net	SAR- Capella Space (50cm/px) Optical- Maxar's WorldView 2 satellite (50cm/px)	SpaceNet, 2020	1 class: building	U-Net	Buildings were easily detected d with the help of U-Net Network	Only one class is considered and tested on only very high spatial resolution SAR data

2.2 Deep learning approaches applied to SAR images

In 1943, Walter Pitts and Warren McCulloch proposed DL [55], a computer model based on human brain neural networks that mimicked the cognitive process through a combination of algorithms and mathematics dubbed 'threshold logic'. Since then many deep network models have been applied on images for various application such as object detection, classification, semantic and instance segmentation, image recognition, visual art processing, image restoration. Model-based (such as physical and statistical models) and data-learning (such as deep neural networks) techniques have recently been applied to SAR imagery to drive diverse algorithms. With the advancement of CNN (Convolutional Neural Network) [56] and state-of-the-art deep learning methodologies, systems may learn complicated land cover properties from a variety of remote sensing images more efficiently. Table 2 illustrates the recent studies where deep/machine learning techniques applied on various SAR image formats using the annotated data hich is mentioned in Section 2.1. All these studies are based on land cover mapping and most of them used Sentinel-1 SAR image as input image. Mehra et al. [46] used multi-temporal single polarization data of VV band and found PSP network best for classification and evaluate the models using IoU technique whereas Gargiulo et al. [52] found U-Net model to be more reliable for classifying the land cover using Sentinel-1 dual-pol (VV and VH) image. Sentinel-1 dual-pol was also used in the paper [49] where 93% overall classification accuracy was achieved by using SegNet and FCDenseNet deep learning models. Authors of paper [48] and [38] fused Sentinel-1 and 2 image for storing textural and spatial information respectively and applied U-Net techniques, Adrian et al. [48] used Sentinel SAR-optical fusion for crop type mapping also included soil, urban and water labels and achived more than 99 percent accuracy using 3D U-Net and Solorzano et al. [38] got good results by applying U-Net model for land cover classification. Gargiulo et al. [54] used VH pol, NDVI and NDWI indices as a 3 band input and proposed W-Net network for land cover mapping which showed good classification result scoring 88% of precision. In the paper [51] [53] [37] author used airborne SAR image with less than 2m pixel resolution as mentioned in Table **??** and applied machine and deep learning for classifying land cover type, Since SAR image has a reasonable spatial resolution, land cover features can be improved and increased for this investigation, even though the model performed well in terms of its application. The majority of the studies used a single polarisation band or a fusion of Sentinel-1 and Sentinel-2 images, which will be difficult to categorise during monsoon or overcast weather. Our research will primarily focus on exploiting Sentinel-1 SAR images without any optical image fusion and developing pre-trained SAR image classification classifiers.

3 Proposed Methodology

Figure 2 represents the schematic diagram of the research. Sentinel-1 SAR data must be pre-processed before being fed to any machine or deep neural networks. The snappy module in Python or Brockmaan Consult's open source Sentinel Application Platform (SNAP) software can be used to conduct pre-processing. SLC (Single Look Complex) and GRD (Ground Range Detected) are the two product types commonly referred to as L1 in Sentinel images, and we used GRD type taken in

IW acquisition mode for our area of interest (AOI) because these data have already been detected and multilooked, and they are projected to ground range using the earth ellipsoid model. But, several prepossessing steps are needed to generate covariance matrix in SLC product [57].

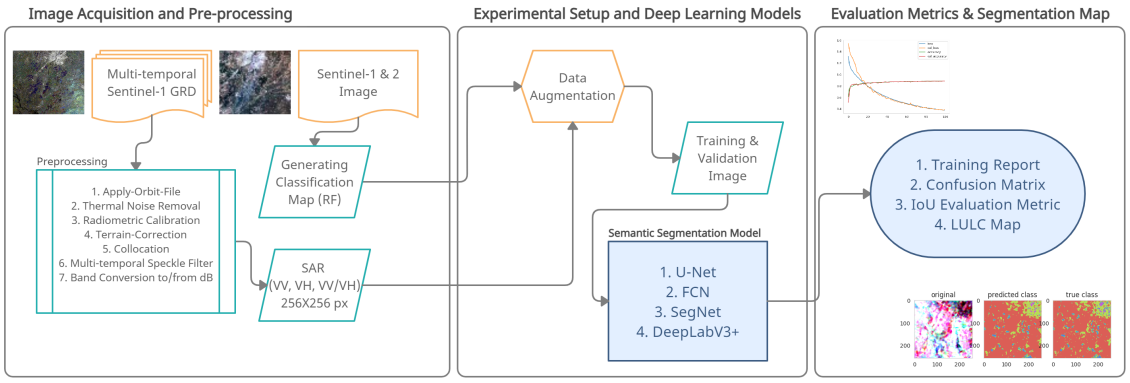

Figure 2: Schematic diagram of the research

3.1 Study Area

Good number of dataset is required for training any deep neural networks, so keeping that in consideration we have taken parts of Ahmedabad, Anand, Kheda region in Gujarat, India, As per Figure 3, this areas have mixed land cover features like flooded vegetation, barren land, built areas, water bodies, agriculture, grassland, forest etc.

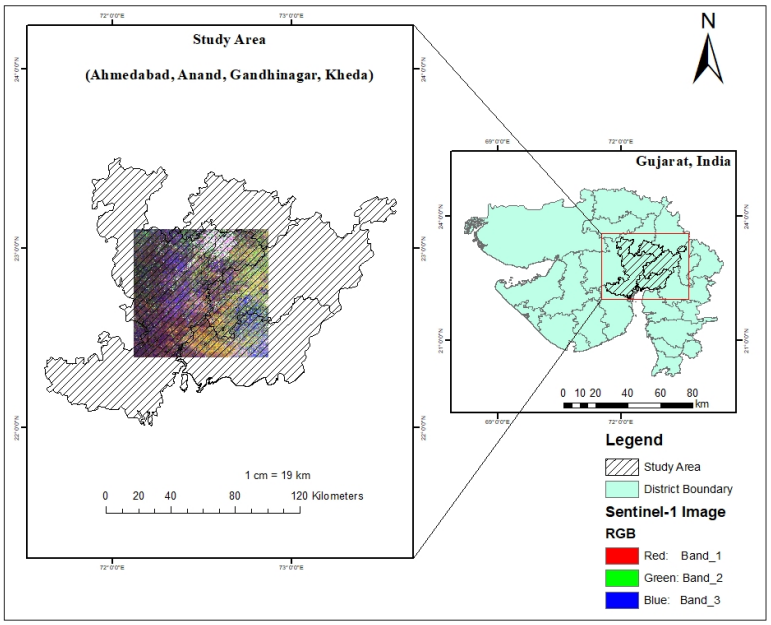

Figure 3: Study Area

3.2 Data Source

Sentinel-1's principal acquisition mode over land is the Interferometric Wide (IW) swath mode. It collects data over a 250-kilometer swath with a spatial resolution of 5m x 20m (single look). Image used for this study had VV, VH quad-pol polarization, spatial resolution of 5*20 metre (ground range by azimuth). Refer Table 3.

Table 3: Acquisition parameters of Sentinel-1 SAR image

Sensing Date	Acquisition Mode	Polarization	Resolution	Wavelength	Pass
13/02/2018	IW	VH, VV	5*20 metre	C-band (5.6 cm)	Descending
07/07/2018	IW	VH, VV	5*20 metre	C-band (5.6 cm)	Descending

3.3 Data Annotating

The training and validation samples for the five land cover features viz., barren land, water, agriculture, build up, dense vegetation were generated from Sentinel-1 image on SNAP software by referring Google Earth image. Google Earth images of 2018 were used to collect training samples for all five land cover features. This data was gathered using QGIS v3.16, an open source desktop geographic information system (GIS) application and applied random forest algorithm using SNAP software. The groundtruth samples were trained on Sentinel-1 data using random forest classification algorithm, with the parameters set to 5000 training samples and 100 trees. On the test dataset, the overall accuracy achieved was 82.46 percent. The resulting groundtruth labels were utilized to create a annotated dataset for overlaying and training on Sentinel-1 SAR data with deep neural network segmentation models. Table 4 shows the classification accuracy of the five landcover features generated using random forest technique.

Table 4: Random Forest performance on Sentinel-1 data using SNAP 8.0

Landcover Type	Accuracy
Agriculture	0.73
Barren Land	0.93
Build up	0.79
Dense vegetation	0.73
Water	0.92
Overall Accuracy	0.82
RMSE	0.96

3.4 Pre-processing of Sentinel-1 image

To pre-process any Sentinel-1 data, there are few steps: apply orbit file, thermal noise removal, border noise removal, radiometric calibration, speckle filtering, terrain correction, and so forth [58].This Pre-processing steps were applied on data sets taken on February 13th and July 07th, 2018, respectively,

in our research. We have used Lee Sigma filter [59] with window size 7*7, sigma 0.9, targeted window size 3*3 in multi-temporal speckle filtering to further compress the speckle noise done after radiometric calibration, the spatial/pixel resolution were re-sampled to 10m, all these methods were executed in SNAP tool [60]. For the proposed model 256X256X3 (height*width*band channel) input image size were chosen for training the five land-cover features (barren land, water, agriculture, dense vegetation, build up) on UNET, FCN, SegNet and DeeplabV3+ segmentation models.

3.5 Experimental Setup

This section will cover the development of training and validation of Sentinel-1 SAR images, image augmentation and implementation of the software and hardware setup.

3.5.1 Development of Training and Validation Data

Same size of image patches is considered for all the five segmentation models i.e., 256 X 256 pixel approximately $2X2km^2$ which is in square shape. Although some models like FCN, Segnet are independent of image size and shape but we have considered same image size to equally compare the performance of the segmentation models.
After pre-processing of Sentinel-1 SAR image as described in section 4.1, two sigma nought (σ_0) bands i.e., Sigma_VH (single co-polarization, vertical transmit/horizontal receive) and Sigma_VV (single co-polarization, vertical transmit/vertical receive) were generated and both were converted into decibel (dB) . For training purposes, these dB values were manually transformed back to intensity values by taking antilog and distributing them across multiple orders of magnitude. The first two channels were represented by the backscatter amplitude for polarizations: VH and VV. VH is assigned to the R channel, whereas VV is assigned to the G channel. As the third channel B, ratio of VH over VV (cross-pol) was included. Overall 900 image patches of 256X256 pixel were generated in which 85 percent of data were used for training the segmentation models and rest 15 percent of data were used for validating the model performance.

3.5.2 Image Augmentation

As mentioned above, 900 image patches were generated from the study area which are quite less for training any deep neural network [61]. To overcome from this, augmentation techniques were applied. The fundamental goal of data augmentation is to increase learning by enhancing original images with little changes like altering the orientation, transposing, adding noise in the image, or adjusting the intensity significantly. This gives the model more information, and the images size are essentially raised. Furthermore, the data augmentation aids the model in learning some unique data qualities for which no image samples are available in the original images [62].
Apart from adding of noise, contrast and other filters in the images, rest of the augmentation parameters (transposing, flipping, rotation) were applied in this research. Since, SAR backscattered values can be altered by adding noises and contrast [63] which can lead to misclassification of landcover features by the segmentation models.

3.5.3 Implementation of Software and Hardware

All the four models were executed on Python 3.6 using Tensorflow backend [64] with keras interface. Below listed are the parameters used to train the SAR images for all four segmentation models:

- Adam optimisation technique with a learning rate of 0.001 and exponential decay rate for first moment estimations of 0.9, and for the second-moment estimates of 0.99.
- We used the early stopping criterion, which meant that after 10 epochs of no improvement in the development (validation) loss, the training would immediately cease for each model.
- Because of the early stopping conditions, different models were trained for varying numbers of epochs, ranging from 69 (for DeepLabV3+) to 126 (for SegNet).
- Before stopping in each scenario, the checkpoint for the most recent model with the best result was kept.
- Then the generated model is used on test dataset to make predictions, and reported the findings.

3.5.4 Model Parameters

For any object-oriented image analysis, segmentation is important [65]. Semantic segmentation assigns class to each pixels on whatever image object or location it relates with, these techniques not just to recognize and locate items in an image, but also provide precise regions and limits for those things. Several segmentation models have been applied on SAR data such as U-Net, Deeplab V3+, fully convolutional network (fcn), SegNet, DenseNet, HCapsNet, BiSeNet etc., ([46],[66],[67]).

U-Net

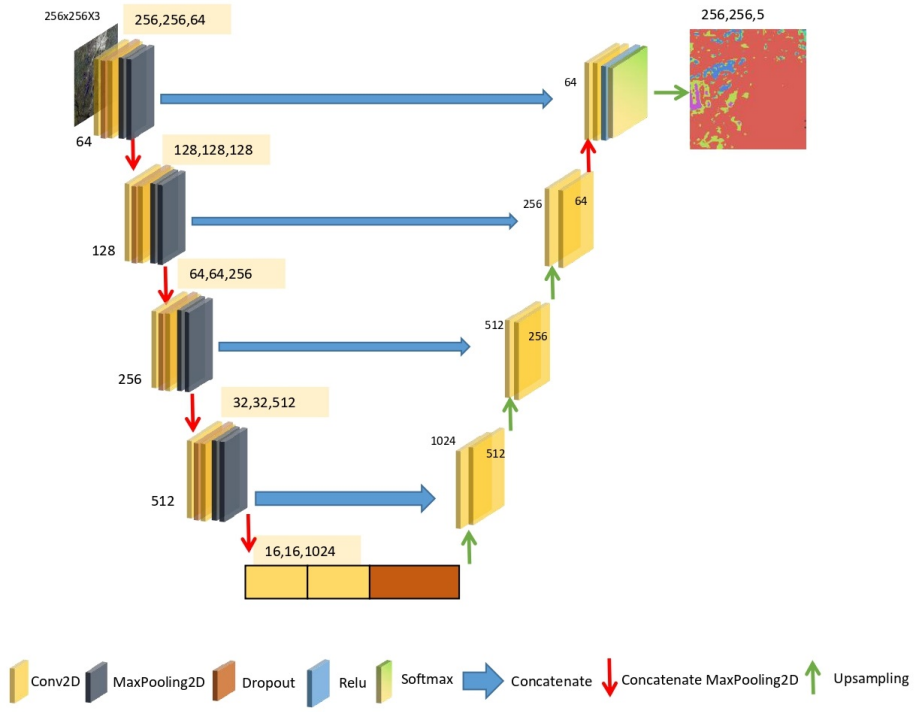

Figure 4: U-Net Architecture [31]

The convolutional network's contracting path follows the standard architecture, consisting of two 3x3 convolutions (same padded convolutions) that are applied repeatedly for three times, each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling (see Figure 4). With each downsampling step, the number of feature channels is quadrupled. A 2x2 convolution (transposed 2D convolutional layer) that halves the number of feature channels, a concatenation with the proportionately cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU in the expansive path follow an upsampling of the feature map.Image segmentation performance has been measured by comparing two segmentation masks: a reference mask with ground truth objects that depict true segmentation vs. a predicted/estimated segmentation mask. Cropping is essential in every convolution. To convert each 64-component feature vector to the desired number of land cover features, a 1x1 convolution is used at the last layer, followed by batch normalisation and the softmax function. There are a total of 23 convolutional layers in the network.

Fully Convolution Network (FCN)

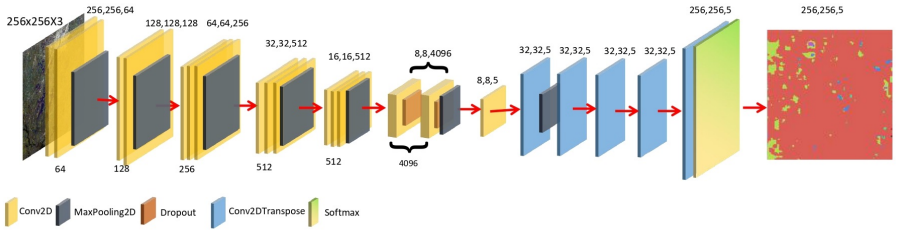

Figure 5: FCN Architecture [36]

Figure 8 illustrate the FCN-8 architecture used in this research, FCN-8 uses the output of VGG-16 (Block 5) and , then concatenates it with the output of Block 4. It then concatenates the Block 3 output with a second upscale x2. Finally, a higher-end x8 is created [68].the model has downsmapling and upsampling network, the first half of the network model involves downsampling which has five blocks of convolutional networks. Block one and two has two 3x3 convolution respectively while block three to five has three 3x3 convolution followed by rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling the image's spatial resolution and creating complex feature mappings. We obtain finer visual information with each convolution. At this point, we have highly effective discrimination across classes; but, the location information has been lost. Downsampling is followed by an upsampling technique to retrieve the position information, which accepts many lower resolution images as input and outputs a high-resolution segmentation map.

SegNet

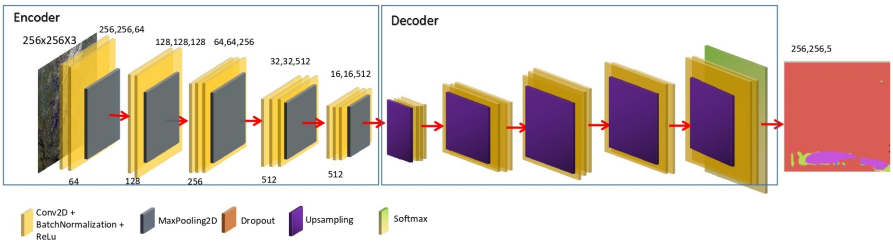

Figure 6: SegNet Architecture [41]

Like UNet and FCN-8, Segnet also have encoder and decoder network part. Figure 6 shows the Segnet architecture used for this research. The encoder part have 13 convolutional layers from VGG-16 skipping the fully connected network, the network also include batch normalization, rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2. Decoder network prduces desired land cover feature maps by upsampling of size 2*2 and convolutional layer followed by batch normalization, kernel initilizaer and with 'same' padding methods. Finally, each pixel's class is predicted using a K-class softmax function.

DeepLabV3+

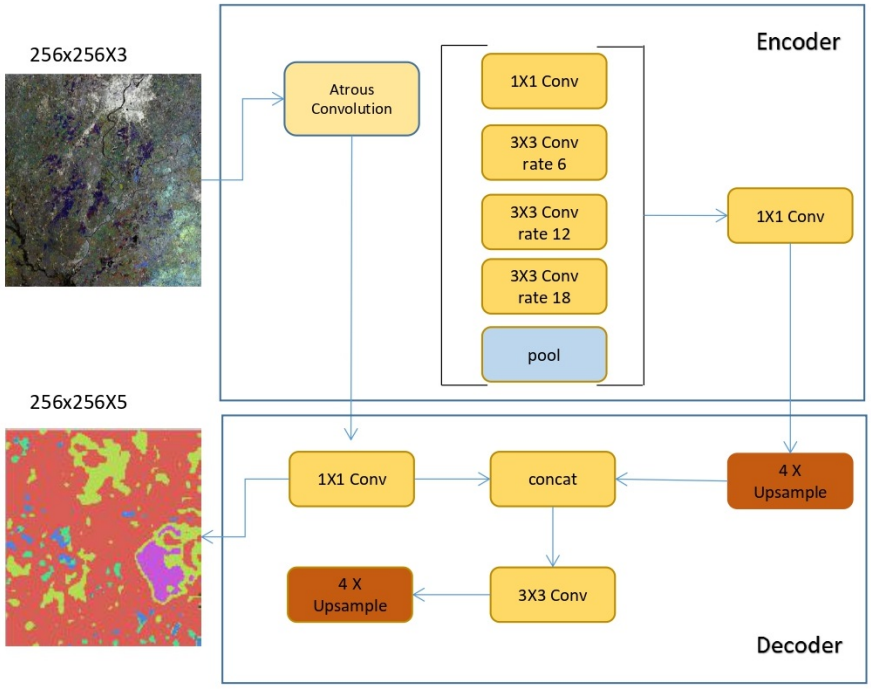

Figure 7: DeepLabV3+ Architecture [43]

Applied atrous convolutional network in the encoder part to extract features at an arbitrary resolution as shown in Figure 7. In the decoder section 4-fold bilinear up-sampling on the ASPP outputs, 1X1 convolutional with reduced filter number on a intermediate feature layer, concatenated ASPP outputs with intermediate features, then applied 3X3 convlutions followed by 4-fold bilinear up-sampling. This decoder will improve the resolution of the image.

3.5.5 Accuracy Metric

The difference between two segmentation masks, a reference mask with ground truth labels indicating actual segmentation against the predicted segmentation label image, has been used to evaluate the pixel-wise segmentation accuracy. Some of the common accuracy assessment techniques used for evaluating computer vision models are root mean square error (RMSE) or mean absolute error (MSE) [69], intersection over union (IoU)/ jaccard index [70], dice coefficient (F1-score) [71] etc. In this research, IoU also known as jaccard index is used to evaluate and compare the performance of the four segmentation models which is a more accurate measure of accuracy than the traditional overall percentage measure and is used to evaluate the correctness of various classes [46]. IoU is the ratio of area of overlap over area of union see Equation 1, area of overlap is the true positive (TP) rate represents the correctly classified pixels while area of union is sum of the true positive (TP),

false positive (FP) and false negative (FN) [72]. Confusion matrix and overall pixel accuracy for training and validation data were also generated to evaluate the performance of the models.

$$IoU = \frac{Area of Overlap}{Area of Union} = \frac{TP}{TP + FP + FN} \tag{1}$$

4 Results and Discussion

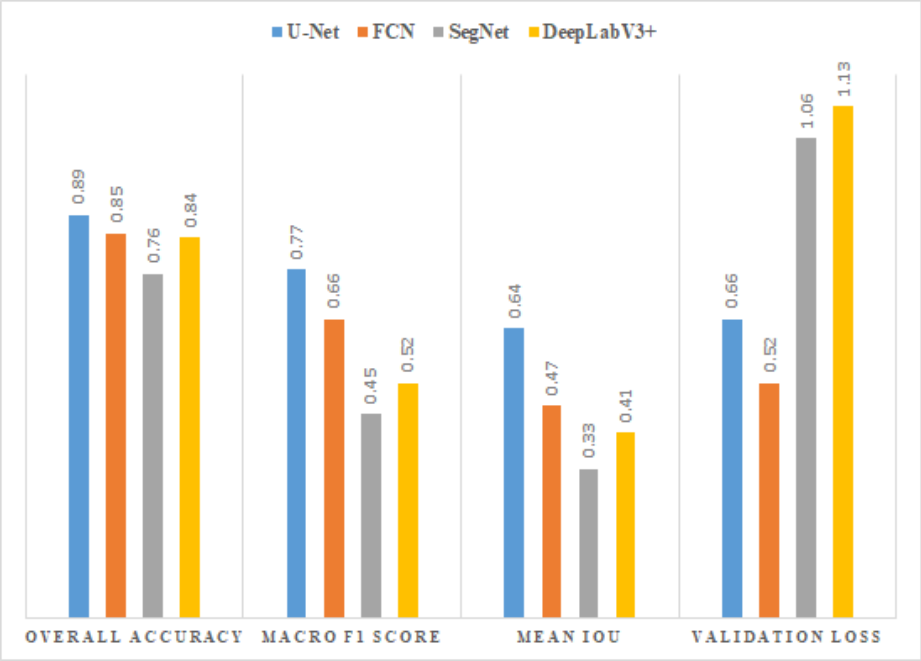

Figure 8: Overall model performance

Table 5 summarizes the four segmentation models performance applied on Sentinel-1 SAR data with the polarization combination of VV, VH and VV ratio VH. The performance were evaluated by comparing the individual IoU and Mean IoU of all the five classes, accuracy and loss function of training and validation, precision, recall, f1 score and confusion matrix of all the four models. Out of four models, U-Net achieved encouraging results of 85% validation accuracy with a loss function of 0.66 and a mean IOU of 0.65. IoU for individual classes in U-Net models as shown in Table 4 led to the observation that Agriculture land features has higher IoU (0.91) followed by water(0.67), Dense vegetation (0.62), and barren land (0.58) . The selected region of interest was more dominated by vegetation and water bodies, which may be one of the reasons for the low IoU value in built-up areas (0.41). Apart from SegNet, all other models performed well. Figure 8 compares the overall accuracy, macro f1 score, mean IoU, and validation loss of four models. Except for the validation loss, U-Net performed well in the remaining three criteria. FCN has a minimal validation loss, whereas SegNet has low accuracy across the board in all four metrics.

Table 7 illustrates the confusion matrix for all four models, with Agriculture having the highest class-wise accuracy of more than 96% for all four models, and U-Net performing well for the remaining classes. SegNet has exhibited very low classification accuracy for barren area, build up, vegetation, and water, and most of the classes are incorrectly classified as agriculture due to vibrant objectives and high errant traits. The reason for the lower scores in build up, barren land, and dense vegetation could be due to feature mixing, such as vegetation mixing with dense agriculture land, rough surface of barren land mixing with build up regions, and vice versa. The results obtained are quite similar to those obtained in the previous research reported in section 2 with different class formation of 3-4 or more core categories and scoring greater than 90% accuracy with fusion of optical and SAR images. When compared to the results of the article [46], which used four land cover characteristics (urban, water, forest, and agricultural) and achieved the mean IoU of 0.64 and 0.61, respectively, utilising the pyramid scene parsing network (PSP) and the U-Net model. Despite the fact that they did not include barren land as a primary land cover feature and used a Landsat-8 satellite image (30m spatial resolution) to create the reference map, our findings showed that increasing the primary class features i.e., including barren land produced better overall results.

Table 5: Segmentation model performances on multi-temporal Sentinel-1 SAR data

LULC Class	***IoU***			
	U-Net	**FCN**	**SegNet**	**DeepLabV3+**
Agriculture	**0.92**	0.86	0.77	0.87
Barren Land	**0.58**	0.38	0.21	0.41
Build Up	**0.41**	0.42	0.12	0.31
Dense Vegetation	**0.62**	0.24	0.16	0.50
Water	**0.67**	0.43	0.38	0.56
Mean IoU	**0.64**	0.47	0.33	0.53
Training Accuracy	0.85	0.79	0.75	0.83
Training Loss	0.66	0.57	0.76	0.44
Validation Accuracy	0.85	079	0.71	0.82
Validation Loss	0.66	0.52	1.06	0.48

Table 6: Segmentation model training report

Train Report												
	U-Net			FCN			SegNet			DeepLabV3+		
	Precision	Recall	f1-score	Precision	Recall	f1-score	Precision	Recall	f1-score	Precision	Recall	f1-score
Agriculture	0.96	0.96	**0.96**	0.9	0.96	0.93	0.78	0.98	0.87	0.90	0.97	0.93
Barren Land	0.74	0.73	**0.73**	0.75	0.53	0.62	0.64	0.23	0.34	0.72	0.48	0..58
Build Up	0.57	0.59	**0.58**	0.46	0.4	0.43	0.64	0.13	0.21	0.57	0.39	0.47
Dense Vegetation	0.75	0.78	**0.76**	0.65	0.65	0.65	0.45	0.2	0.28	0.61	0.72	0.66
Water	0.74	0.88	**0.80**	0.64	0.72	0.67	0.63	0.49	0.55	0.69	0.74	0.72
accuracy			**0.89**			0.85			0.76			0.85
macro avg	0.75	0.79	**0.77**	0.68	0.65	0.66	0.63	0.41	0.45	0.70	0.66	0.67
weighted avg	0.89	0.89	**0.89**	0.84	0.85	0.84	0.73	0.76	0.71	0.84	0.85	0.84

Table 7: Confusion Matrix of four segmentation models

U-Net					
Actual / Predicted	**Agriculture**	**Barren Land**	**Build Up**	**Dense Vegetation**	**Water**
Agriculture	**0.96**	0.03	0.01	0.00	0.00
Barren Land	0.16	**0.73**	0.09	0.02	0.00
Build Up	0.06	0.22	**0.59**	0.12	0.01
Dense Vegetation	0.01	0.06	0.13	**0.78**	0.02
Water	0.00	0.12	0.00	0.00	**0.88**
FCN					
Agriculture	**0.96**	0.02	0.01	0.00	0.00
Barren Land	0.35	**0.53**	0.08	0.03	0.01
Build Up	0.32	0.11	**0.40**	0.16	0.01
Dense Vegetation	0.11	0.05	0.17	**0.65**	0.01
Water	0.00	0.25	0.00	0.03	**0.72**
SegNet					
Agriculture	**0.98**	0.02	0.00	0.02	0.00
Barren Land	0.71	**0.23**	0.02	0.03	0.01
Build Up	0.77	0.02	**0.13**	0.07	0.00
Dense Vegetation	0.74	0.02	0.03	**0.20**	0.01
Water	0.17	0.31	0.00	0.02	**0.49**
DeepLabV3+					
Agriculture	**0.97**	0.02	0.01	0.00	0.00
Barren Land	0.40	**0.48**	0.07	0.05	0.01
Build Up	0.30	0.12	**0.39**	0.18	0.01
Dense Vegetation	0.13	0.06	0.08	**0.72**	0.01
Water	0.01	0.24	0.00	0.01	**0.74**

5 Conclusion and Future Scope

The use of predefined deep neural networks for segmentation on SAR images has further improved the capabilities to generate detailed land cover mapping. Use of time-series data on SAR images captures

variability in the feature class and further enhance the training accuracy when compared to single time-frame image. Findings in this research demonstrate the potential for applying state-of-the-art semantic segmentation deep learning models to Sentinel-1 multi-temporal image classification. From this research, main land cover features like water, agriculture land, build up areas and barren land can be easily classified with dual-pol Sentinel-1 SAR data and further classes can be increased with the availability of Quad-pol SAR data in future studies. Compared to model-based approaches, learning algorithms that take advantage of an advanced processing framework and a spate of relevant SAR data are more flexible, robust to data, and efficient, resulting in improved performance. Yet, when confined to a small data collection, complicated situations, and other circumstances, the learning algorithms may suffer from poor generalisation, low feature detection resilience, and impracticality. More research is needed to construct new theories, starting with models, concepts, and architecture designs, in order to boost the growth of machine and deep learning in SAR imagery.

References

[1] J. R. Anderson, *A land use and land cover classification system for use with remote sensor data*. US Government Printing Office, 1976, vol. 964.

[2] N. Kumar, S. Yamaç, and A. Velmurugan, "Applications of remote sensing and GIS in natural resource management," *Journal of the Andaman Science Association*, vol. 20, no. 1, pp. 1–6, 2015.

[3] H. K. Kija, J. O. Ogutu, L. J. Mangewa, J. Bukombe, F. Verones, B. J. Graae, J. R. Kideghesho, M. Y. Said, and E. F. Nzunda, "Spatio-temporal changes in wildlife habitat quality in the greater Serengeti ecosystem," *Sustainability*, vol. 12, no. 6, p. 2440, 2020.

[4] Y. Wu, S. Li, and S. Yu, "Monitoring urban expansion and its effects on land use and land cover changes in Guangzhou city, China," *Environmental monitoring and assessment*, vol. 188, no. 1, p. 54, 2016.

[5] N. Y. D. Twumasi, Z. Shao, and A. Orhan, "Remote sensing and GIS methods in urban disaster monitoring and management–an overview," *International Journal of Trend in Scientific Research and Development*, vol. 3, no. 4, pp. 918–926, 2019.

[6] Y. Xie, W. Dai, Z. Hu, Y. Liu, C. Li, and X. Pu, "A novel convolutional neural network architecture for SAR target recognition," *Journal of Sensors*, vol. 2019, 2019.

[7] M. Chamling and B. Bera, "Spatio-temporal patterns of land use/land cover change in the Bhutan–Bengal foothill region between 1987 and 2019: study towards geospatial applications and policy making," *Earth Systems and Environment*, pp. 1–14, 2020.

[8] K. Koehn, K. Brye, and C. Scarlat, "Quantification of stormwater runoff using a combined GIS and curve number approach: a case study for an urban watershed in the Ozark Highlands, USA," *Urban water journal*, vol. 8, no. 4, pp. 255–265, 2011.

[9] S. F. Fonji and G. N. Taff, "Using satellite data to monitor land-use land-cover change in North-eastern Latvia," *Springerplus*, vol. 3, no. 1, pp. 1–15, 2014.

[10] J. Spruce, J. Bolten, I. N. Mohammed, R. Srinivasan, and V. Lakshmi, "Mapping land use land cover change in the Lower Mekong Basin from 1997 to 2010," *Frontiers in environmental science*, vol. 8, p. 21, 2020.

[11] S. Abdikan, F. B. Sanli, M. Ustuner, and F. Calò, "Land cover mapping using sentinel-1 SAR data," in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B7, 2016 XXIII ISPRS Congress*, 2014.

[12] I.-H. Holobâcă, K. Ivan, and M. Alexe, "Extracting built-up areas from Sentinel-1 imagery using land-cover classification and texture analysis," *International Journal of Remote Sensing*, vol. 40, no. 20, pp. 8054–8069, 2019.

[13] C. Kontgis, M. S. Warren, S. W. Skillman, R. Chartrand, and D. I. Moody, "Leveraging Sentinel-1 time-series data for mapping agricultural land cover and land use in the tropics," in *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, IEEE, 2017, pp. 1–4.

[14] E. Attema, P. Bargellini, P. Edwards, G. Levrini, S. Lokas, L. Moeller, B. Rosich-Tell, P. Secchi, R. Torres, M. Davidson, *et al.*, "Sentinel-1-the radar mission for GMES operational land and sea services," *ESA bulletin*, vol. 131, pp. 10–17, 2007.

[15] R. Torres, P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, B. Rommen, N. Floury, M. Brown, *et al.*, "GMES Sentinel-1 mission," *Remote Sensing of Environment*, vol. 120, pp. 9–24, 2012.

[16] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geoscience and remote sensing magazine*, vol. 1, no. 1, pp. 6–43, 2013.

[17] A. Veloso, S. Mermoz, A. Bouvet, T. Le Toan, M. Planells, J.-F. Dejoux, and E. Ceschia, "Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications," *Remote sensing of environment*, vol. 199, pp. 415–426, 2017.

[18] P. W. Vachon, J. Wolfe, and H. Greidanus, "Analysis of Sentinel-1 marine applications potential," in *2012 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2012, pp. 1734–1737.

[19] M. El Hajj, N. Baghdadi, M. Zribi, and S. Angelliaume, "Analysis of Sentinel-1 radiometric stability and quality for land surface applications," *Remote Sensing*, vol. 8, no. 5, p. 406, 2016.

[20] A. Barra, O. Monserrat, P. Mazzanti, C. Esposito, M. Crosetto, and G. Scarascia Mugnozza, "First insights on the potential of Sentinel-1 for landslides detection," *Geomatics, Natural Hazards and Risk*, vol. 7, no. 6, pp. 1874–1883, 2016.

[21] S. Bousbih, M. Zribi, Z. Lili-Chabaane, N. Baghdadi, M. El Hajj, Q. Gao, and B. Mougenot, "Potential of Sentinel-1 radar data for the assessment of soil and cereal cover parameters," *Sensors*, vol. 17, no. 11, p. 2617, 2017.

[22] G. Forkuor, J.-B. B. Zoungrana, K. Dimobe, B. Ouattara, K. P. Vadrevu, and J. E. Tondoh, "Above-ground biomass mapping in West African dryland forest using Sentinel-1 and 2 datasets-A case study," *Remote Sensing of Environment*, vol. 236, p. 111 496, 2020.

[23] X. Zhu, S. Montazeri, M. Ali, Y. Hua, Y. Wang, L. Mou, Y. Shi, F. Xu, and R. Bamler, "Deep learning meets SAR: concepts, models, pitfalls, and perspectives," *IEEE Geoscience and Remote Sensing Magazine (GRSM)*, 2021.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[25] G. Joseph and S. Dhawan, *Fundamentals of remote sensing. Hyderabad*, 2005.

[26] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[27] S. Du, S. Du, B. Liu, and X. Zhang, "Incorporating DeepLabv3+ and object-based image analysis for semantic segmentation of very high resolution remote sensing images," *International Journal of Digital Earth*, vol. 14, no. 3, pp. 357–378, 2021.

[28] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[29] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Bisenet: Bilateral segmentation network for real-time semantic segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 325–341.

[30] T. Pohlen, A. Hermans, M. Mathias, and B. Leibe, "Full-resolution residual networks for semantic segmentation in street scenes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4151–4160.

[31] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.

[32] Y. Li, L. Xu, J. Rao, L. Guo, Z. Yan, and S. Jin, "A Y-Net deep learning method for road segmentation using high-resolution visible remote sensing images," *Remote sensing letters*, vol. 10, no. 4, pp. 381–390, 2019.

[33] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, Ieee, 2017, pp. 1–6.

[34] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," *arXiv preprint arXiv:1409.2329*, 2014.

[35] P. Liang, W. Shi, and X. Zhang, "Remote sensing image classification based on stacked denoising autoencoder," *Remote Sensing*, vol. 10, no. 1, p. 16, 2018.

[36] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[37] R. A. Emek and N. Demir, "Building detection from sar images using unet deep learning method," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 44, pp. 215–218, 2020.

[38] J. V. Solórzano, J. F. Mas, Y. Gao, and J. A. Gallardo-Cruz, "Land Use Land Cover Classification with U-Net: Advantages of Combining Sentinel-1 and Sentinel-2 Imagery," *Remote Sensing*, vol. 13, no. 18, p. 3600, 2021.

[39] K. A. Korznikov, D. E. Kislov, J. Altman, J. Doležal, A. S. Vozmishcheva, and P. V. Krestov, "Using U-Net-Like Deep Convolutional Neural Networks for Precise Tree Recognition in Very High Resolution RGB (Red, Green, Blue) Satellite Images," *Forests*, vol. 12, no. 1, p. 66, 2021.

[40] Y. Guo, X. Cao, B. Liu, and M. Gao, "Cloud detection for satellite imagery using attention-based U-Net convolutional neural network," *Symmetry*, vol. 12, no. 6, p. 1056, 2020.

[41] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition. arXiv," *arXiv preprint arXiv:1409.1556*, 2018.

[42] V. Badrinarayanan, A. Handa, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling," *arXiv preprint arXiv:1505.07293*, 2015.

[43] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[44] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[45] C. Bansal, H. O. Ahlawat, M. Jain, O. Prakash, S. A. Mehta, D. Singh, H. Baheti, S. Singh, and A. Seth, "IndiaSat: A Pixel-Level Dataset for Land-Cover Classification on Three Satellite Systems-Landsat-7, Landsat-8, and Sentinel-2," in *ACM SIGCAS Conference on Computing and Sustainable Societies*, 2021, pp. 147–155.

[46] A. Mehra, N. Jain, and H. S. Srivastava, "A novel approach to use semantic segmentation based deep learning networks to classify multi-temporal SAR data," *Geocarto International*, pp. 1–16, 2020.

[47] M. Delalay, V. Tiwari, A. D. Ziegler, V. Gopal, and P. Passy, "Land-use and land-cover classification using Sentinel-2 data and machine-learning algorithms: operational method and its implementation for a mountainous area of Nepal," *Journal of Applied Remote Sensing*, vol. 13, no. 1, p. 014 530, 2019.

[48] J. Adrian, V. Sagan, and M. Maimaitijiang, "Sentinel SAR-optical fusion for crop type mapping using deep learning and Google Earth Engine," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 175, pp. 215–235, 2021.

[49] S. Šćepanović, O. Antropov, P. Laurila, Y. Rauste, V. Ignatenko, and J. Praks, "Wide-Area Land Cover Mapping with Sentinel-1 Imagery using Deep Learning Semantic Segmentation Models," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 10 357–10 374, 2021.

[50] G. Büttner, "CORINE land cover and land cover change products," in *Land use and land cover mapping in Europe*, Springer, 2014, pp. 55–74.

[51] R. Garg, A. Kumar, N. Bansal, M. Prateek, and S. Kumar, "Semantic segmentation of PolSAR image data using advanced deep learning model," *Scientific reports*, vol. 11, no. 1, pp. 1–18, 2021.

[52] M. Gargiulo, D. A. Dell'Aglio, A. Iodice, D. Riccio, and G. Ruello, "Semantic Segmentation using Deep Learning: A case of study in Albufera Park, Valencia," in *2019 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, IEEE, 2019, pp. 134–138.

[53] X. Wang, L. Cavigelli, M. Eggimann, M. Magno, and L. Benini, "Hr-SAR-NET: A deep neural network for urban scene segmentation from high-resolution SAR data," in *2020 IEEE Sensors Applications Symposium (SAS)*, IEEE, 2020, pp. 1–6.

[54] M. Gargiulo, D. A. Dell'Aglio, A. Iodice, D. Riccio, and G. Ruello, "Integration of sentinel-1 and sentinel-2 data for land cover mapping using W-Net," *Sensors*, vol. 20, no. 10, p. 2969, 2020.

[55] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[56] Y. LeCun, Y. Bengio, *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[57] D. Mandal, D. S. Vaka, N. R. Bhogapurapu, V. Vanama, V. Kumar, Y. S. Rao, and A. Bhattacharya, "Sentinel-1 SLC preprocessing workflow for polarimetric applications: a generic practice for generating dual-pol covariance matrix elements in SNAP S-1 toolbox," 2019.

[58] F. Filipponi, "Sentinel-1 GRD preprocessing workflow," in *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 18, 2019, p. 11.

[59] J.-S. Lee, J.-H. Wen, T. L. Ainsworth, K.-S. Chen, and A. J. Chen, "Improved sigma filter for speckle filtering of SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 202–213, 2008.

[60] M. Zuhlke, N. Fomferra, C. Brockmann, M. Peters, L. Veci, J. Malik, and P. Regner, "SNAP (sentinel application platform) and the ESA sentinel 3 toolbox," in *Sentinel-3 for Science Workshop*, vol. 734, 2015, p. 21.

[61] D. J. Benkendorf and C. P. Hawkins, "Effects of sample size and network depth on a deep learning approach to species distribution modeling," *Ecological Informatics*, vol. 60, p. 101 137, 2020.

[62] S. Scepanovic, O. Antropov, P. Laurila, V. Ignatenko, and J. Praks, "Wide-area land cover mapping with Sentinel-1 imagery using deep learning semantic segmentation models," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, p. 10 357, 2021.

[63] S. Chauhan and H. S. Srivastava, "Comparative evaluation of the sensitivity of multi-polarized SAR and optical data for various land cover classes," *Int. J. Adv. Remote Sens. GIS Geogr*, vol. 4, no. 1, pp. 1–14, 2016.

[64] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, 2016, pp. 265–283.

[65] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS Journal of photogrammetry and remote sensing*, vol. 80, pp. 91–106, 2013.

[66] S. Scepanovic, O. Antropov, P. Laurila, V. Ignatenko, and J. Praks, "Wide-area land cover mapping with Sentinel-1 imagery using deep learning semantic segmentation models," 2019.

[67] J. Cheng, F. Zhang, D. Xiang, Q. Yin, Y. Zhou, and W. Wang, "PolSAR Image Land Cover Classification Based on Hierarchical Capsule Network," *Remote Sensing*, vol. 13, no. 16, p. 3132, 2021.

[68] S. A. Kamran and A. S. Sabbir, "Efficient yet deep convolutional neural networks for semantic segmentation," in *2018 International Symposium on Advanced Intelligent Informatics (SAIN)*, IEEE, 2018, pp. 123–130.

[69] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)," *Geoscientific Model Development Discussions*, vol. 7, no. 1, pp. 1525–1534, 2014.

[70] R. Shi, K. N. Ngan, and S. Li, "Jaccard index compensation for object segmentation evaluation," in *2014 IEEE international conference on image processing (ICIP)*, IEEE, 2014, pp. 4457–4461.

[71] A. Fujino, H. Isozaki, and J. Suzuki, "Multi-label text categorization with model combination based on f1-score maximization," in *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II*, 2008.

[72] A. Luque, A. Carrasco, A. Martın, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, vol. 91, pp. 216–231, 2019.