Pixel-Based Classification of Major Rice Varieties using Sentinel-2 L2A Data via CNN in Spectral and Time Domain

Author

Muhammad Usman Rauf

00000172306

Supervisor

Dr. Hamid Jabbar

DEPARTMENT OF MECHATRONICS ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

AUGUST, 2020

Pixel-Based Classification of Major Rice Varieties using Sentinel-2 L2A Data via CNN in Spectral and Time Domain

Author

Muhammad Usman Rauf

00000172306

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Mechatronics Engineering

Supervisor:

Dr. Hamid Jabbar

Supervisor's Signature:_____

DEPARTMENT OF MECHATRONICS ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

AUGUST, 2020

Declaration

I certify that this research work titled "*Pixel-Based Classification of Major Rice Varieties using Sentinel-2 L2A Data via CNN in Spectral and Time Domain"* is my own work. The work has not been presented elsewhere for evaluation. The material that has been used from other sources it has been properly acknowledged/referred.

Signature of Student

Muhammad Usman Rauf

MS-MTS-00000172306

Language Correctness Certificate

This thesis has been read by an English expert and is free of typing, syntax, semantic, grammatical, and spelling mistakes. The thesis is also according to the format given by the university.

Signature of Student

Muhammad Usman Rauf

MS-MTS-00000172306

Signature of Supervisor

Copyright Statement

Acknowledgments

I am thankful to my creator **Allah Subhanahu-Wa Ta'ala** to have directed me throughout this work at every step and for every new thought which You put in my mind to improve it. Indeed I could have done nothing without Your priceless help and guidance. Whosoever helped me throughout my thesis, whether my parents or any other individual was Your will, so indeed none be worthy of praise but You.

I am generously thankful to my beloved ***parents*** who raised me when I was not capable of walking and continued to support me throughout every stage of my life.

I would like to special thanks to my supervisor **Dr. Hamid Jabbar** and co-supervisor **Dr. Waqar Shahid Qureshi** for their enormous support and assistance. Each time I got stuck in something, they came up with the solution. Without their help, I wouldn't have been able to complete my thesis. I would also like to thank **Dr. Uzair Khaleeq-uz-Zaman** and **Dr. Danish Hussain** for being on my thesis guidance and evaluation committee and also thankful to **Dr. Umar Shahbaz** and **Dr. Mubashir Saleem** for their support and cooperation for course work which they have taught me.

I would also like to pay thanks to my colleagues especially **Dr. Raheel Qureshi** for facilitating me whenever I required his support and guidance throughout the whole thesis.

Finally, I would like to express my gratitude to all the individuals who have rendered valuable assistance to my study.

“Dedicated to my exceptional parents and adored siblings whose enormous support and assistance led me to this wonderful achievement.”

Abstract

Crop monitoring and management, efficient utilization of natural resources, and development of sustainable systems have an important role in keeping agriculture efficacy of crops with high productivity. However, in the area of crop monitoring, it remains a challenge to monitor crop varieties due to the lack of accurate data available on time. In Pakistan, rice crop is important both agriculturally and economically. Two major rice varieties including Basmati and IRRI were selected to classify in the present study.

Satellite remote sensing offered a suitable approach as the bases for coverage, accuracy, cost-effectiveness, well revisit time, and type of data which hold great potential for routine applications in crop monitoring and management. Sentinel-2A/B on board Multispectral Instrument (MSI) delivers data with a superior spectral, spatial, and temporal resolution, makes them suitable for crop monitoring. Deep learning techniques such as convolutional neural networks (CNN) in agriculture outperforms the existing approaches by extract the distinguishable and representative features of different land cover from remote sensing images in a hierarchical way to classify. In the present study, a pixel-based deep CNN is devised in both spectral and time domains to map rice varieties in Punjab province, Pakistan. We proposed a spectral unmixing based feature extraction strategy, which provides sub-pixel level information and determines the pixel's abundance map before supervised classification. Abundance representation allows identifying refined structures that are not available in the ground truth map. Vegetation and water indices are also adopted over the full rice-growing season to increase the classification accuracy. Experimental results exhibited an excellent overall accuracy of 98.6 % with the proposed approach.

Key Words: Remote Sensing, Sentinel-2, Deep Learning, Convolutional Neural Network, Rice Crop Classification, Vegetation Indices, Spectral unmixing.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

1.1 Contextual Background

Food is crucial for human survival. About 10000 to 15000 years ago humans began to mold nature for their need and started to cultivate the land in many places around our planet and feed the estimated population of 6 to 10 million at that time. So we must be good enough at it by now. But in current and future scenarios, the world's population is expected to reach around 9 billion in 2030 and 10 billion in 2050. Also, the environmental concerns related to agriculture such as climate change, deforestation, soil degradation, irrigation problems, air quality, and agricultural practices around the world make it more challenging than ever [1]. To keep up the pace there is a strong requirement of timely and efficient monitoring and management of crops to eradicate the problems. In this context, remote sensing technology by its frequent, accurate, timely, cost-effective, and dynamic information can provide efficient monitoring and management and also help in understanding the impact of environmental variables. Different methods can be used to gather this information but satellite remote sensing offers a suitable approach based on the coverage, revisit frequency, and type of data that are provided.

Satellite remote sensing has been used in agriculture since the launch of Landsat-1 in 1972 by NASA [2]. Crop types or varieties mapping are also essential in monitoring, management, and decision making such as market forecast and food security issues. Satellites multispectral and multitemporal images are used in identifying crop types and their phenology stages. Landsat and MODIS satellites multispectral imagery is most frequently used as a data source. However, both have limitations in terms of spatial (MODIS) and temporal resolution (Landsat) [3] [4] [5]. In this context, the European Space Agency (ESA) started the Copernicus program which aimed to increase the spatial, spectral, and temporal resolution by launching twin platform Sentinel-2A and 2B [6]. The Sentinel-2 open data policy with higher spatial resolution, increased number of spectral bands, and high revisit time make it of great interest to the agricultural community and researchers worldwide.

Sentinel-2A and Sentinel-2B were launched in June 2015 and in March 2017 respectively. Both satellites revolving around the earth with a phase angle of 180º in a sun-synchronous orbit at an altitude of 786 km. Sentinel-2 covers the land surface between latitude 56º south to 80º north

including coastal waters 20 km from shore and an island with an area larger than 100 km^2 [7]. Each Sentinel-2 satellite carries a single optical imaging payload called MultiSpectral Instrument (MSI), that provides global coverage with temporal resolution of 2 ~ 5 days depending on latitude, wide view angle or swath of 295 km, 10 to 60 m spatial resolution, and spectral resolution of 13 bands which cover visible, near-infrared (NIR), and shortwave infrared (SWIR) region. Sentinel-2 with these features overthrow the issues with coarse, less frequent, and expensive satellite imagery. In Pakistan, where the crop monitoring activities are usually carried out by conventional methods that are less accurate, time-consuming, and costly, the Sentinel-2 imagery system has great potential to make sure the on-time availability of accurate and unbiased data of crop type or varieties monitoring and classification.

Along with the high-quality data of remote sensing, intelligent and advanced algorithms are necessary to obtain more accuracy and precision in the classification of crop varieties. In past years traditional crop classification methods (knowledge-based [8] and phenology-based [9] [10]), and conventional classification methods such as Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT) methods are successfully used in crop classification. In the last few years, deep learning algorithms like Convolutional Neural Network (CNN) are used in crop type and species classification and outperformed other approaches. The traditional method of classification needs domain knowledge and field surveys for feature mining that can be varied with a change in the dataset. It would cause an obstacle for the availability of data on timely for decision making. On the other hand conventional methods also requires significant domain knowledge for the extraction of featues from raw data. But deep learning algorithms provides hierarchical repesentation of data which allows to address full problem and better learning capabilities. CNN as revolutionary algorithm outperformed other methods in terms of accuracy using both spectral and temporal features of the area of interest.

1.2 Problem Statement and Inspiration

In Pakistan, agriculture is the largest sector in terms of the labor force, and the majority of the population’s livelihood is directly or indirectly depends on it. Also in Pakistan's economy, the agriculture sector is still considered as the backbone, and contribution to GDP is about 19.3 % but there is a lot of potential in that sector to not only for the domestic population but to have excess production for export, which ensure food security as well as increase its share in GDP by utilizing

the latest technologies. Farmers in Pakistan grow their crops in a very challenging like climate change, pest attack, disease outbreak, shortage of water, and market inflation. Pakistan's government realizing the challenges and potential in the agriculture sector and start Prime Minister Agriculture Emergency Program to help farmers, efficient food production that contributes towards food security and sustainable system development [11].

Therefore crop monitoring and management are very important in the development of effective sustainable systems. In developing countries such as Pakistan, crop information is collected through field surveys, which require a lot of human resources, costly, time-consuming, and are not accurate. In these circumstances, satellite remote sensing imagery motivates to research in the field of crop classification and mapping which is the first step towards crop monitoring and management, also an effort in the development of sustainable agriculture systems. Additionally, in remote sensing, the traditional and conventional machine learning methods are not effective as compared to deep learning techniques, which also adds motivation to conduct this research work. Rice crop is selected for this research because it holds an important place in the agriculture sector as it is used as the second main staple food crop after wheat and major exportable product after cotton. Also, rice crop is known as water-thirsty crop and use a lot of resources in addition to that in Pakistan the traditional methods of cultivation waste a lot of natural resources that need to be efficiently managed using new and upcoming state of the art technologies [12].

1.3 Research Aim and Objective

The present work intends to support the rice crop classification and mapping using Sentinel-2 level 2A data in the Multan district of Punjab Pakistan with a deep learning algorithm. To fulfill that following are the objectives:

1. Review the existing deep learning approaches utilized for rice crop classification and select suitable architecture.
2. Extract the sub-pixel level details from ground truth data using spectral unmixing methods.
3. Extract the rice crop growing phenology details from vegetation indices.
4. Implement the chosen architecture and optimized its performance.
5. Evaluate the performance of classification using performance metrics and compare it with the existing work.

1.4 Thesis Organization

The thesis is split up into seven chapters. The first chapter gives an introduction, contextual background, problem statement, and motivation behind that work, research work aim, and objectives. The second chapter gives the theoretical background of spectroscopy, satellite remote sensing, and its characteristics, spectral unmixing, Sentinel-2 mission details, and characteristics, and in last about convolutional neural network. The third chapter provides the literature review of the previous work with a special focus on rice crop classification. The review is divided into three portions according to the approaches used, the first section about traditional, second conventional machine learning, and in third deep learning, approaches discussed. Also, the deep learning approaches are divided into two subsections one is with using ground truth map and second using an abundance map. The fourth chapter presents the detail about the study area, the collection of ground truth data, the satellite imagery dataset, and the preparation of the dataset. The fifth chapter describes the methodology, experimental settings, training and validation process, and evaluation of the CNN architecture using performance metrics. In the sixth chapter the results are shown and discussion on it. In the seventh and last chapter, the thesis ends with a conclusion and future work

2. THEORETICAL BACKGROUND

This chapter will give a theoretical groundwork and start with a short explanation of the field of spectroscopy. After that in the following three sections gives details about satellite-based imaging spectroscopy, its characteristics and in last Sentinel-2 mission was discussed. As the CNN algorithm has been chosen for this study, the second section will give background concepts and detailed architecture of CNN.

2.1 Spectroscopy

In the rainbow, the spectral nature of light is present, but its significance was not recognized until in the 17th century Isaac Newton describe the colors of the rainbow that combine to make white light of the sun. In 1672 Isaac Newton presented the word "*spectrum*" to explain this phenomenon by conducting a series of experiments and demonstrated that the sunlight could be divided into distinct bands of colors [13], [14]. Newton's experiments of light were the opening of the science of spectroscopy. Since than spectroscopy plays a significant role in physics, chemistry, and astronomy. After that, at the beginning of the 18th century, Newton's model was improved and. W Herschel and J.W. Ritter have been proved that the sunlight has span beyond the visible region from infrared to ultraviolet region. In the 1860s Bunsen and Kirchhoff established a link between chemical composition and their unique spectral patterns. They demonstrated that spectroscopy could be used to trace elements or compounds by their unique spectrum. Now after these advancements, spectroscopy became a scientific discipline and used as a tool for analyzing the material's chemical composition [15].

2.1.1 Imaging Spectroscopy or Spectral Imaging

Electromagnetic radiation was the first energy source used in spectroscopic studies and the type of radiative energy involved distinguished the types of spectroscopy. Imaging spectroscopy is an extension of spectroscopy where the light is recorded as images. Spectral imaging may use the ultraviolet, visible, near-infrared, mid-infrared, thermal infrared, microwave, and some other combinations from spectrum also known as spectral bands that can extract additional information that a human vision fails to capture. In spectral imaging for each pixel in an image, it is possible to capture hundreds of wavelength bands. In Multispectral imaging typically 3 to 15 spectral bands

are used to capture radiations but in Hyperspectral imaging tenth to hundreds of contiguous spectral bands are available.

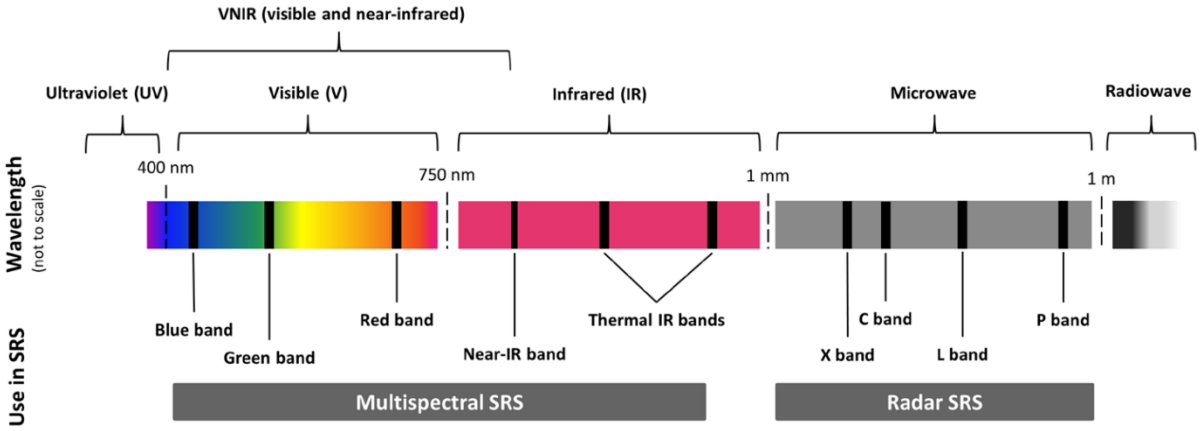

***Figure 1**: Electromagnetic Spectrum. Source: [16]*

An imaging spectrometer is used to obtain a spectral image of an area of interest or scene, often represented as a data cube due to the three-dimensional representation. Two are the image horizontal and vertical distance axes and the third axis is for wavelength.

2.1.2 Space-based Imaging Spectrometer

The imaging spectrometers are also used to observe the features of planet Earth which include vegetation type and health condition assessment, mineral detection for potential mining, and the assessment of water pollution in various water resources by utilizing spectral data obtained from satellites. [17]. Prism spectrometers are suitable for Earth observation because of wide spectral ranges typically from 400 nm to 2,500 nm. Hyperion was the first imaging spectrometer to routinely acquire data of the planet earth from space [18]. With the development of imaging spectrometers, several satellites with multispectral imaging sensors were launched successfully into the earth's orbit by NASA and ESA, like Landsat-1~8, PROBA, PRISM, Sentinel-1~3 for gathering earth's surface information from space also referred as satellite remote sensing. Satellite remote sensing used in many fields such as ecology, geography, land surveying, hydrology, and resource monitoring as it gives unique information about earth' surface. To better understand what

is satellite remote sensing? And the terms associated with it A brief overview is given in the next section.

2.2 Satellite Remote Sensing

In satellite remote sensing onboard special sensors are used to obtain information about the features of the earth's surface (soil, water, rocks, and vegetation) without making physical contact with it. The information is gathered by illuminating the object or area of interest by a source of energy and measuring the reflected or emitted radiations from the objects and area. The amount of energy reflected depending upon the properties of the material (chemical, physical, and structural) of the object which helps in detection and discriminating it [19]. So, the source of energy to illuminate the object is a prerequisite in remote sensing. Remote sensing is also categorized based on the energy source. Two broad categories are active remote sensing and the other is passive remote sensing. **Active** remote sensors have their own source of energy and produce electromagnetic radiations and either the amount of energy reflected or the return time is measured by the sensor such as RADAR and light detection and ranging (LiDAR). The main advantage of these sensors over the passive is that they can operate day or night and also under cloudy conditions. On the other hand, **passive** sensors use the natural source of electromagnetic energy (sun) and measure the refected amount of energy from the earth's surface. Typical passive sensors are photographic camera and electro-optical sensors. Passive remote sensors can only operate in the day time and are also affected by cloudy weather conditions [20], [21]. Our main concern is on passive sensors because of the scope of the thesis.

2.2.1 Electromagnetic Radiation and Spectrum

Electromagnetic radiation is a form of energy that propagates through space and medium as a wave-like motion at a speed of light. The electromagnetic wave consisted of two perpendicular components, oscillating electric and magnetic fields (**Figure 2**).

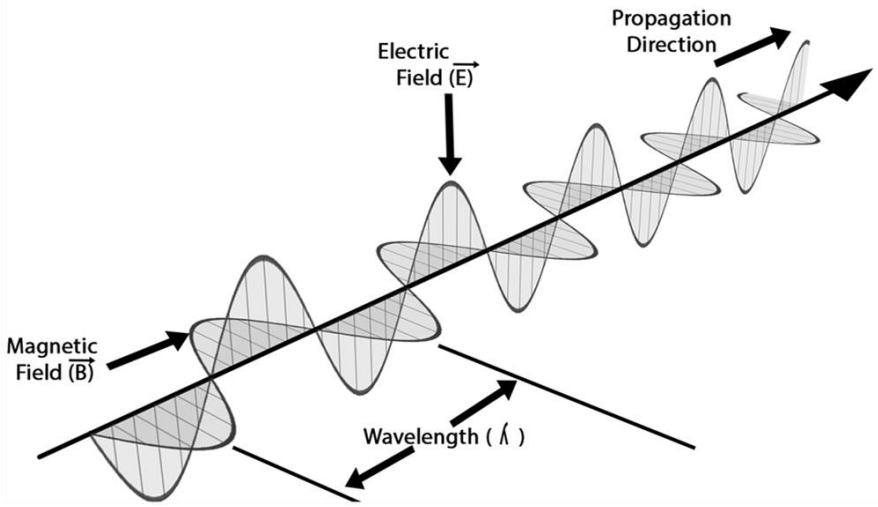

***Figure 2**: Electromagnetic wave. Source:[22]*

Two parameters that characterized the electromagnetic wave are wavelength and frequency and calculated using equation $c = \lambda\nu$. Electromagnetic waves with different wavelength/frequency have different characteristics and practical applications. The continuum of all the electromagnetic waves according to frequency and wavelength is known as the electromagnetic spectrum. The spectrum ranging from one hertz to 10^{25} hertz and divided into separate bands. Each band of frequency is called with a different name such as radio waves, microwaves, infrared, ultraviolet, visible, X-rays, etc (Figure 3).

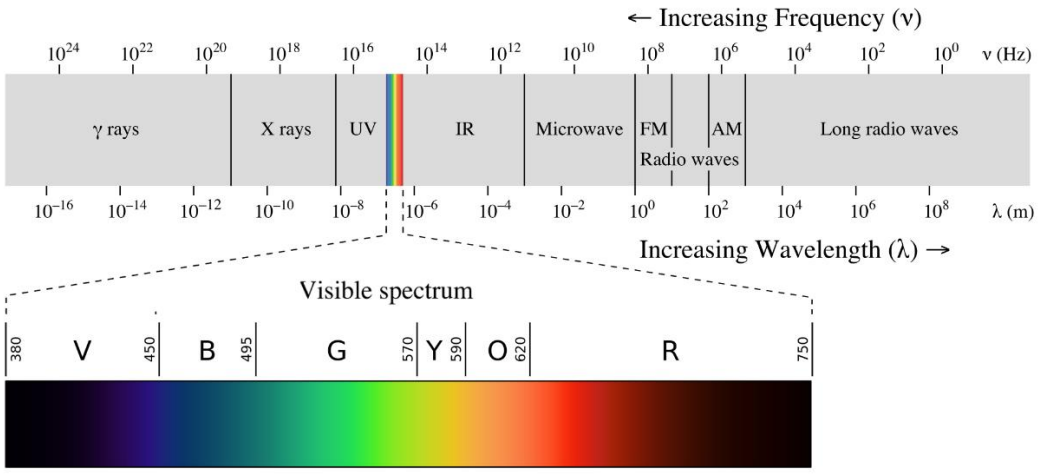

***Figure 3**: Electromagnetic Spectrum. Source: [23]*

2.2.2 Interaction with Atmosphere

The electromagnetic radiations from the sun pass through the earth's atmosphere twice, once from the sun to the earth's surface and second after being reflected from the earth's surface to the satellite sensor. Atmosphere interaction with direct sunlight and the reflected light affect it due to gases and small particles in the atmosphere and interfere remote sensing process are known as **Atmospheric Effect**. Three physical phenomena, scattering, absorption, and refraction occurred during travel through the atmosphere. **Scattering** redirects the radiations by the suspended particles and gasses in the atmosphere. The extent of scattering is affected by the particle size, abundance, wavelength of radiations, and the distance traveled through the atmosphere. As the concentration of the particles in the atmosphere varies so the scattering effect also uneven spatially and temporally. **Absorption** occurred by the gas molecules in the atmosphere which absorbs the radiations as it passes through it in certain spectrum bands and reduces the amount of light (**Figure 4**).

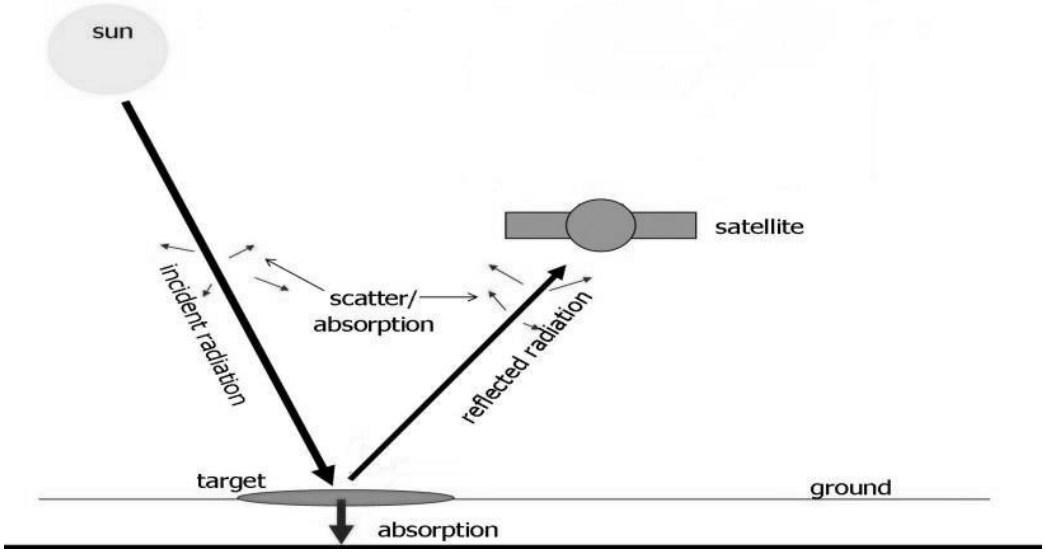

***Figure 4**: Atmospheric effect on electromagnetic radiation from the sun. Source: [24].*

The area of the spectrum bands which are not influenced by atmospheric absorption and hit the earth's surface are called **atmospheric windows** and useful for remote sensing. In remote sensing, sensors are built to choose those specific portions of the electromagnetic spectrum.

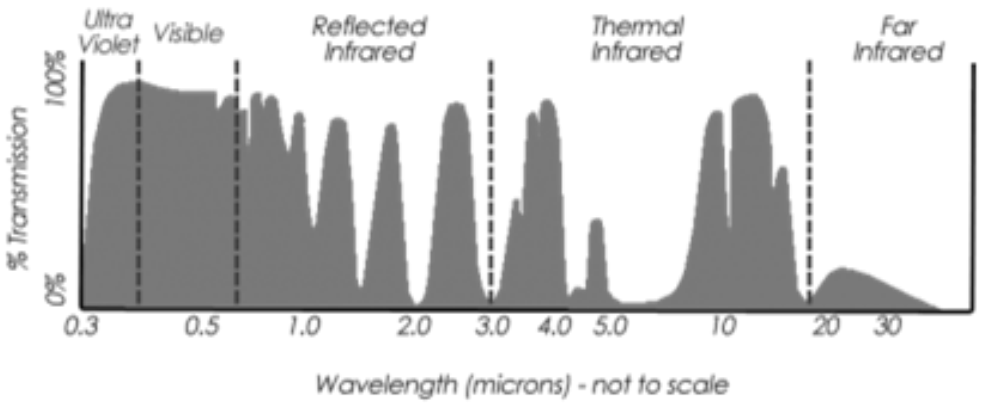

***Figure 5**: Atmospheric Window. Source: [25]*

The atmospheric windows are present in the UV region, almost completely visible region and some portions of the infrared spectrum (**Figure 5**).

2.2.3 Interaction with Earth's Surface

Radiations from the sun that are not scattered or absorbed reached to the earth's surface. On interaction with the earth's surface, several changes occurred in incident radiations that are detected by the remote sensors and obtain useful information. Upon interaction, it is either reflected, transmitted, or absorbed depending upon the wavelength, material, and conditions. **Transmission** occurs when the radiations passed through the object without significant attenuation and **absorption** when the radiations are completely absorbed by it. In remote sensing, the most useful and revealing are the radiations that are bounced back from the object or reflected from it. Two types of reflections are specular reflection and diffused reflection. In **specular reflection,** almost all the radiations are reflected like a mirror because of the smooth surface but on a rough surface **diffused reflection** occurs and the radiations are reflected uniformly in all directions.

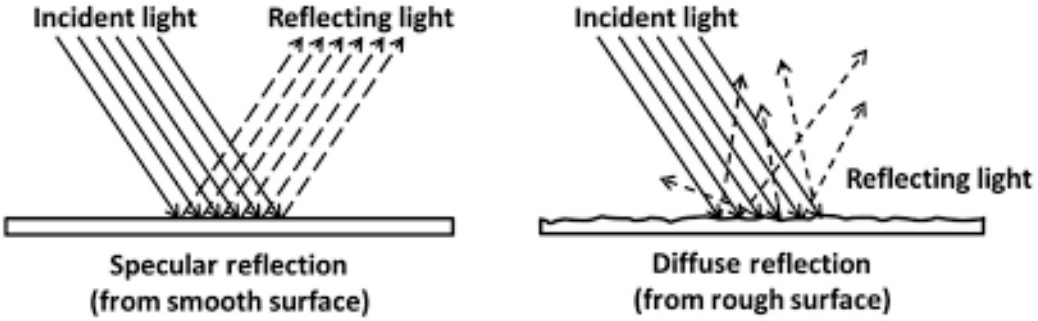

***Figure 6**: Specular and Diffused Reflection. Source: [26]*

2.2.4 Spectral Reflectance

The ratio between reflected energy and incident energy in terms of wavelength is called **spectral reflectance**. Different materials have different spectral reflectance values at specific wavelength intervals comprise the **spectral signatures** which are used to distinguish them. Spectral reflectance depends upon the wavelength and changes with the variations in the object's chemical composition and environmental conditions. The reflectance characteristics of soil, vegetation, and water are shown in **Figure 7**.

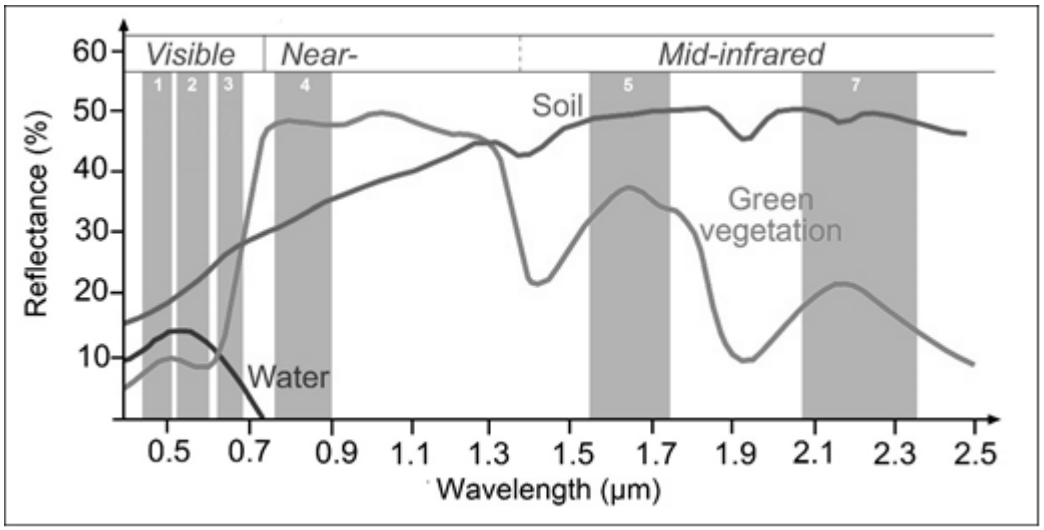

***Figure 7**: Spectral reflectance curve of soil, vegetation, and water. Source: [27]*

The soil reflectance characteristics are determined by its amount of moisture, organic composition, and texture. Soil reflectance curve shows less peak because the majority of the radiations are either reflected or absorbed with very small transmittance. In vegetation, the chlorophyll in the leaves reflect the green wavelength but absorbs the red and blue wavelength. Vegetation health is determined by measuring and monitoring the spectral reflectance values in near-infrared. In the case of water, the majority of the radiations are either absorbed or transmitted, only in blue or blue-green region wavelength are reflected and thus water looks in that color range. These variations depend upon the water depth, materials in water, and the surface of the water body.

Spectral Index

As the earth's features have different spectral signatures, a combination of two or more spectral reflectance from different wavelengths is used to extract an abundance of features of interest is known as a spectral index. These indices are developed to monitor or detect changes in the features

and help to predict and model the phenomenon. Spectral indices highlight the spectral properties, reduce background effect on the feature of interest, and also the effects atmosphere. Vegetation, water, and soil indices are most commonly used to detect, monitor, and extract specific features from them. The most common indices are:

***Table 1**: Spectral Indices and derived formulas.*

Normalized Difference Vegetation Index (NDVI) [28]	$NDVI = \frac{NIR - RED}{NIR + RED}$
Soil Adjusted Vegetation Index (SAVI) [29]	$SAVI = \frac{NIR - RED}{NIR + RED + L} * (1 + L)$
Normalized Difference Water Index (NDWI) [30]	$NDWI = \frac{GREEN - NIR}{GREEN + NIR}$
Normalized Difference Moisture Index (NDMI) [31]	$NDMI = \frac{NIR - SWIR}{NIR + SWIR}$

2.3 Characteristics of Satellite and Sensor

To use data obtained from remote sensing, it is important to understand the characteristics of the platform (satellite) and onboard sensors. The most important characteristics of the satellite are orbit, and data capturing mechanism and for sensor resolutions (spatial, spectral, radiometric), that has to be understood.

2.3.1 Orbit and Swath

The path followed by the satellite is called its orbit and selected according to the onboard sensor capabilities. The selection of orbit varies in terms of altitude, orientation, and rotation around the earth. Two orbits that are important for the observations of the earth are geo-stationary and polar orbit. In a **geostationary** orbit, satellite keeps pace with the earth rotation from west to east at the equatorial plane and covering the same geographical area day and night. The altitude of the geostationary orbit is located at 36000 km above the equator.

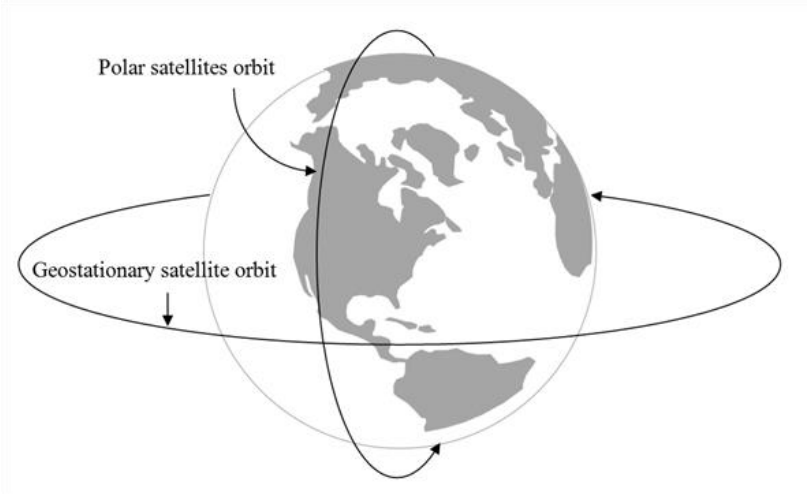

***Figure 8**: Satellite orbits. Source: [32]*

The second orbit which is mostly used in remote sensing is **polar** because of its cycle around the earth from near north to south pole rather than west to east as in geostationary orbit. The altitude of the polar orbit varies from 600 to 900 km approximately. **Sun-synchronous** orbit is a kind of polar orbit in which satellite passes over the polar region and also synchronized with the sun. sun-synchronous means that the satellite always visits the same geographical area at the same local time.

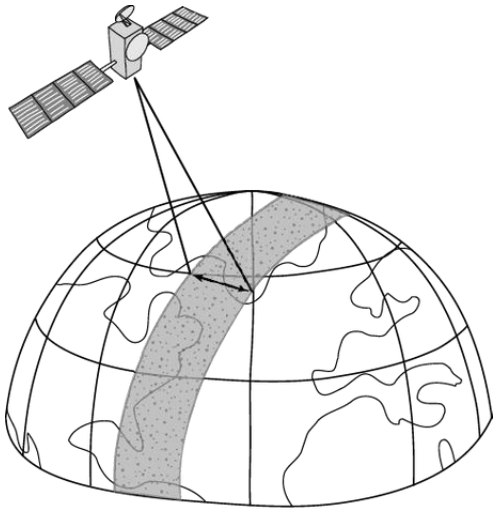

***Figure 9**: Satellite Swath.*

During revolving around the earth satellite capture a certain portion of the earth's surface is referred to as **swath**. The combined rotations of satellite and earth from north to south and west to

east allow to covering a new geographical area with each consecutive pass. The satellite's orbit, swath, and the earth rotation allow the repetitive coverage of the entire earth periodically.

2.3.2 Resolution

The quality of the remote sensing data depends upon the four key parameters of the platform and sensor, which include spatial extant, the electromagnetic spectrum bands, the time interval between acquisition, and radiation quantity.

Spatial Resolution

Spatial resolutions refer to the smallest detectable feature or the area on the ground that each pixel cover. In the fine resolution, the pixel size is smaller and small objects or features can be detected but in low or coarse resolution only large objects or features are detected due to larger pixel size.

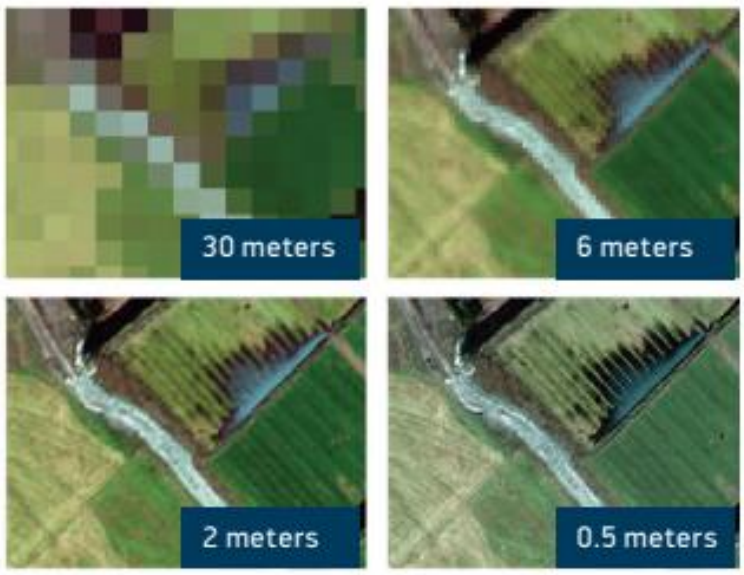

***Figure 10**: Spatial resolution effect on imagery.*

The spatial resolution mainly depends on the instantaneous field of view (IFOV) which is the angular cone of the visibility of the sensor viewed from a given altitude at any instant of time. IFOV, when expressed in terms of ground distance covered by a single pixel, is referred to as the ground instantaneous field of view (GIFOV) and calculated by multiplying the IFOV and altitude from the ground.

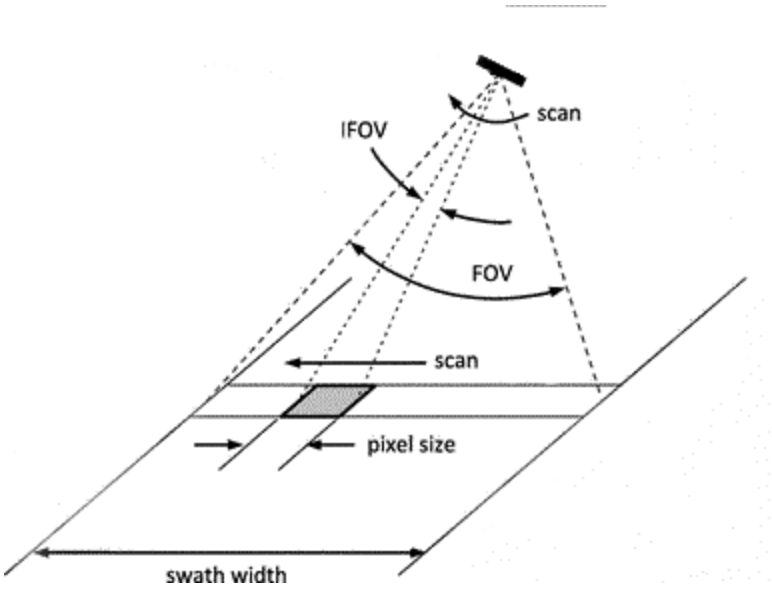

***Figure 11**: Instantaneous Field of View (IFOV).*

Spectral Resolution

The spectral resolution of the sensor can be described by three components: the number of spectral bands, range of wavelength covered by each band, and the location of the band in the electromagnetic spectrum. Multispectral sensors have multiple separate spectral bands in the visible, NIR, SWIR, and thermal infrared regions. In hyperspectral sensors, the number of bands increases but the bandwidth or the range of wavelength of single-band decreased.

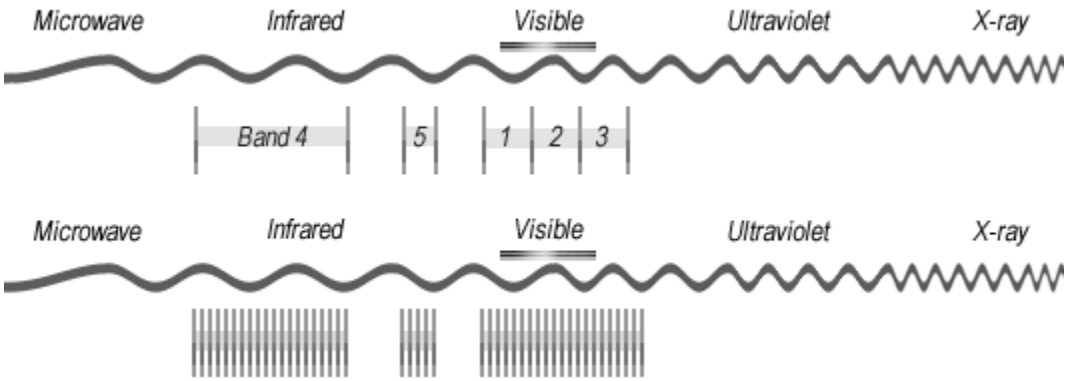

***Figure 12**: The spectral resolution of multispectral & hyperspectral sensors. Source: [25]*

Different features or details of the object can be distinguished by comparing their spectral responses at distinctive bands. Coarse features like water and vegetation can be separated by multispectral sensors, whereas fine details like rock types and tree species required narrow wavelength bands to separate them.

Radiometric Resolution

The radiometric resolution provides information about the sensor that can distinguish a small change in radiation intensity. A sensor that has greater sensitivity in detecting small differences in reflectance has a fine radiometric resolution. Radiometric resolution is represented by positive digital numbers ranging depends on the number of bits used in representing radiation intensity levels. Thus if a sensor uses 8 bits it records 2^8=256 intensity levels from 0 to 255and if 4 bit uses only 16 levels from 0 to 15 are available.

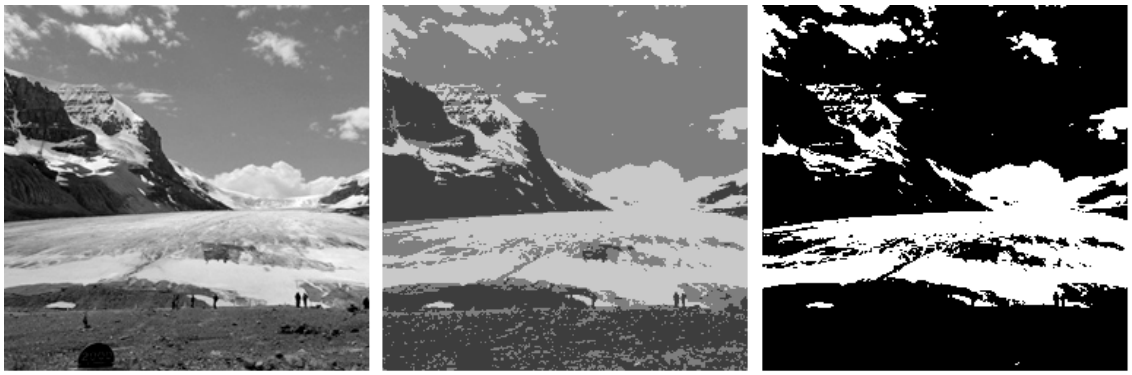

***Figure 13**: Detail of image at different radiometric resolution.*

If we compare an 8-bit image with a 4-bit image, a large difference can be seen in the level of details because of radiometric resolution.

Temporal Resolution

In remote sensing temporal resolution is also an important concept to consider. Temporal resolution is the time in which a satellite return and capture the same geographical area on the surface of the earth also known as return time or revisit time. The temporal resolution depends upon several factors such as satellite orbit, sensor capability, swath, and latitude. Some geographical area is re-imaged more frequently due to overlap of imaging swath, and increasing latitude.

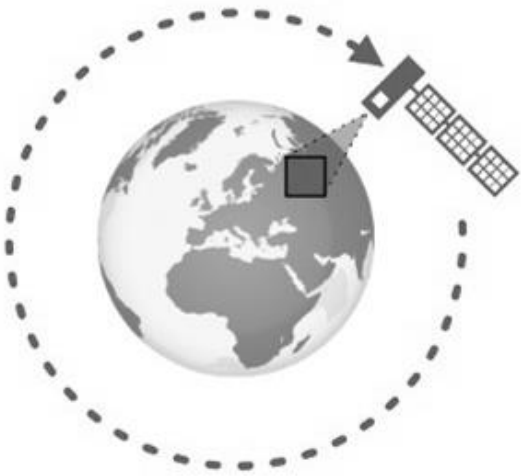

***Figure 14**: Temporal resolution with a single satellite.*

Temporal resolution is increased by using multiple satellites together as a constellation. Multi-temporal imagery is most important in remote sensing as spectral characteristics and features change over time and it helps in detecting these changes. A comparison between different satellite programs is given below in **Table 2** [33].

***Table 2**: Resolution comparison of satellite missions.*

Platform	Resolution			
	Spatial (m)	Spectral (bands)	Temporal (days)	Radiometric (bits)
MODIS	250-1000	36	1-2	12
LandSat-8	15-100	11	16	12
Sentinel-2	10-60	13	2-5	12
Rapid Eye	5	5	1	12
WorldView-3	0.3-30	29	1-5	11-14

2.4 Sentinel-2

In 1998, the European Space Agency (ESA) was started the Global Monitoring for Environment and Security (GEMS) program, nowadays known as Copernicus [34]. Sentinel-2 is also included in the Copernicus program that provides services to monitor global terrestrial surfaces and coastal water. In the next section, some details about the Sentinel-2 satellite and data are provided.

2.4.1 Mission Overview

The complete Sentinel-2 program comprises of two satellites (Sentinel-2A and 2B) orbiting around the earth in the same polar orbit but phased at 180º from each other. Sentinel-2A and Sentinel-2B were launched in June 2015 and March 2017 respectively, in a sun-synchronous orbit at 786 km altitude with a descending node (North to South pole) at 10:30 A.M.

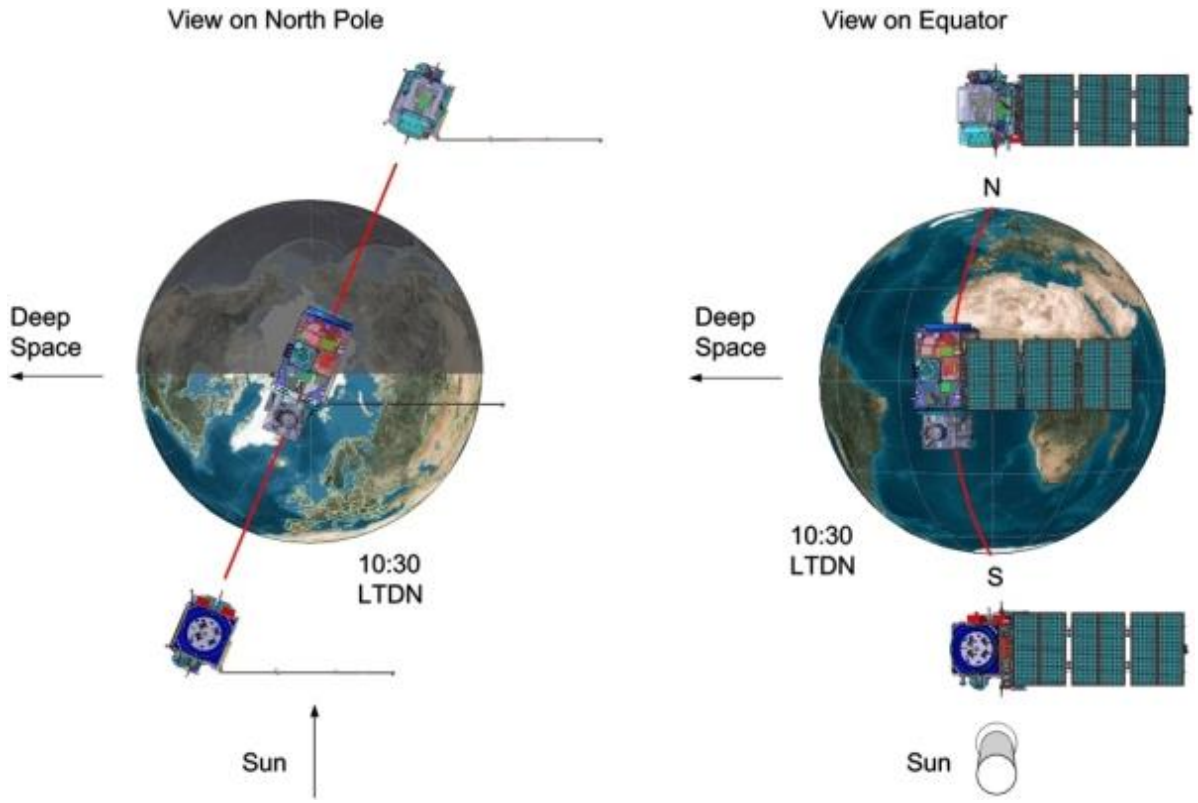

***Figure 15**: Sentinel-2 twin satellite orbital configuration. Source: [35]*

Each Sentinel-2 satellite possesses a single optical imaging payload called MultiSpectral Instrument (MSI). Sentinel-2 covers the land surface between latitude 56º south to 80º north including coastal waters 20 km from shore and an island with an area larger than 100 km^2.

Sentinel-2 key mission objectives are:

- Systematic global acquisition of multispectral images with high resolutions and revisit frequency.
- Continue the multispectral imagery providing by SPOT and Landsat satellites.

- Provides observation data such as land use land cover map, land change detection, and geophysical variables [7].

2.4.2 Sensor Overview

The Sentinel-2 mission with two identical satellites and MSI provides a systematic global coverage with a high temporal resolution of 5 days at the equator and wide view angle or swath of 295 km. The MSI measures the reflected radiance in 13 spectral bands, which cover the visible (VIS), near-infrared (NIR), and shortwave infrared (SWIR) regions in the electromagnetic spectrum with a spatial resolution of 10m to 60 m. MSI has 4 bands with 10 m spatial resolution, Blue (490 nm), Green (560 nm), Red (665 nm), and NIR (842 nm), 6 bands at 20 m spatial resolution, Red Edge (705 nm, 740 nm, 783 nm), Narrow NIR (865 nm), and SWIR (1619 nm, 2190 nm), and 3 bands have 60 m spatial resolution, Aerosols (443 nm), Water vapor (940 nm), and Cirrus (1380 nm) (**Table 3**) [36] [37].

***Table 3**: Sentinel-2 MSI spectral bands, wavelength, bandwidth, and spatial resolution.*

Bands	Use	Sentinel-2A	Sentinel-2B	Bandwidth (nm)	Spatial Resolution (m)
		Central Wavelength (nm)	Central Wavelength (nm)		
B1	Aerosols	442.7	442.2	21	60
B2	Blue	492.4	492.1	66	10
B3	Green	559.8	559.0	36	10
B4	Red	664.6	664.9	31	10
B5	Red Edge 1	704.1	703.8	15 ~ 16	20
B6	Red Edge 2	740.5	739.1	15	20
B7	Red Edge 3	782.8	779.7	20	20
B8	NIR	832.8	832.9	106	10
B8A	Narrow NIR	864.7	864.0	21 ~ 22	20
B9	Water vapor	945.1	943.2	20 ~ 21	60
B10	Cirrus	1373.5	1376.9	30 ~ 31	60
B11	SWIR 1	1613.7	1610.4	91 ~ 94	20
B12	SWIR 2	2202.4	2185.7	175 ~ 185	20

2.4.3 Product Overview

Product Processing Levels

The processing levels of Sentinal-2 MSI are:

- **Level-0 or L0 Product:** Raw sensor data for long time storage.
- **Level-1A or L1A Product:** Uncompressed sensor data with rough registration in sensor geometry without radiometric correction and only used for calibration purposes.
- **Level-1B or L1B Product:** Top-of-Atmosphere (TOA) radiance in sensor geometry with radiometric correction applied, and used for calibration, validation, and quality control purposes.
- **Level-1C or L1C Product:** TOA reflectance in cartographic geometry.
- **Level-2A orL2A Product:** Bottom-of-Atmosphere (BOA) reflectance in cartographic geometry. This product can be generated from the L1C product by using the Sen2Cor processor publically.

The processing levels from level-0 to level-1C are presented in **Figure 16**.

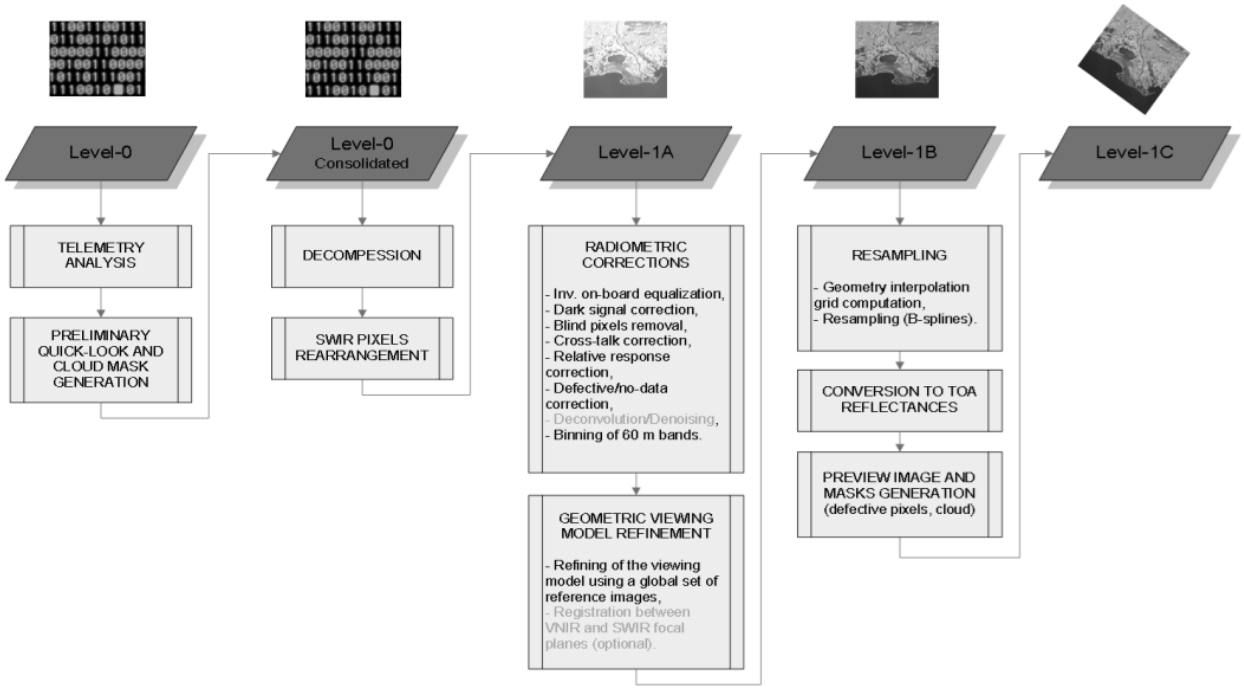

***Figure 16**: Processing levels from Level-0 to Level-1C. Source: [36]*

Form level-0 to level-2A all the systematic processing and archiving is done by the Payload Data Ground Segment (PDGS). Only the level-2A is either generated by PDGS or by the Sen2Cor processor on the user side. The quality of the generated data is evaluated by OnLine Quality Control (OLQC).

Product Type

Two Sentinel-2 products, Level-1C and Level-2A, either generated by ground segment or Sen2Cor toolbox are available for the users. The description of both products are given in

***Table 4**: Sentinel-2 online available product types.*

Product Type	Description	Generation & Distribution	Data Volume
Level-1C	TOA reflectance in cartographic geometry	Systematic generation and online distribution	100x100 km^2 600~700 MB single image
Level-2A	BOA reflectance in cartographic geometry	Systematic generation and using toolbox by the user, and online distribution	100x100 km^2 800~900 MB single image

Level-1C

L1C product provides TOA normalized reflectance for each band, having values ranging from 1 (min. reflectance = 0) to 10000 (max. reflectance = 1), also the value of 0 is for "No Data". Sentinel-2 L1C each tile has a fixed size of 100x100 km^2 ortho-image in Universal Transverse Mercator (UTM) or World Geodetic System 1984 (WGS84) projection. Global Digital Elevation Model (DEM) is used to compute the projection. The UTM system splits the earth in 60 zones and each zone has 8º width in latitude and 6º width in longitude. L1C provides each spectral band image in JPEG2000 and solar zenith angle, viewing angle, sun illumination angle for each spectral band on 5 km^2 grid is also provided as metadata in XML format.

Level-2A

The main purpose of generating the L2A surface reflectance product is to remove the atmospheric effects from the L1A product. Therefore, L2A is generated from the ground segment or by the user

using Sen2Cor using the same tiling geometry and projection as in the L1C product. The main processing steps are given in **Figure 17**.

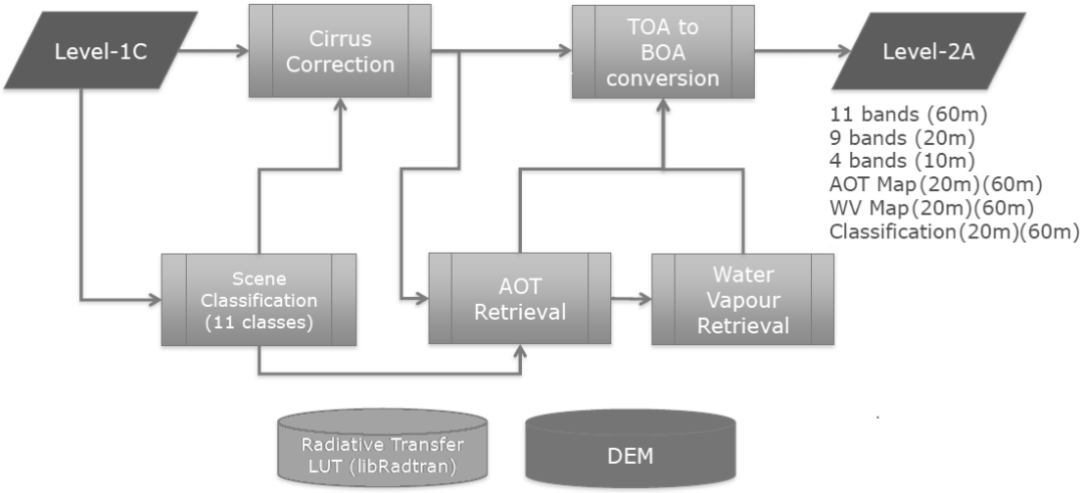

***Figure 17**: Level-2A processing steps. Source: [37]*

Level-2A processing begins with cirrus cloud correction and scene classification (SCL) after that Aerosol Optical Thickness (AOT) and Water Vapor (WV) contents are retrieved from the L1C product. The last step in conversion from TOA to BOA. SCL map is presented in eleven classes and also used as input for atmospheric correction. L2A product also provides additional maps at different spatial resolutions with BOA reflectance images. AOT and WV maps are provided in 10, 20, and 60 m, but cloud, SCL, and snow are only in 20 and 60 m resolutions. The surface reflectance values are also coded in JPEG200 as in Level-1C with 10000 quantization levels. Digital Numbers (DN) are converted into surface reflectance values by using a factor of 1/10000 is applied. In L2A product out of 13 spectral bands, 12 bands are provided, the B10 band at 60 m resolution is excluded because it does not give any information about the surface reflectance and only used for cirrus cloud detection.

2.5 Convolutional Neural Network

Convolutional Neural Network is one type of deep learning network and the most impressive form of Artificial Neural Network architecture today often called CNNs or CovNets. The name also indicates that a mathematical operation named **"convolution"** is applied in the network. Convolution is a special kind of linear operation and applied at least in one of the layers in the network instead of general matrix multiplication which is used in traditional neural networks. CNN

is a biologically inspired network and performs impressively in computer vision problems like object detection, pattern recognition, and especially in image classification and revealed as a state-of-the-art method. CNN's are composed of neurons that optimized it-self by learning.

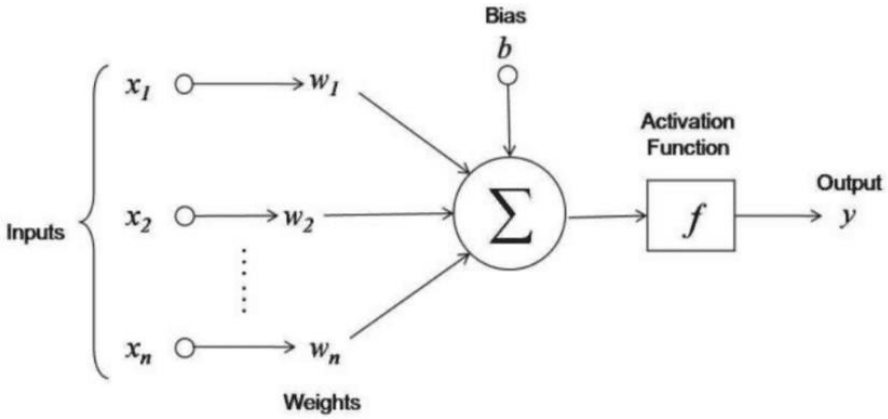

***Figure 18**: Single neuron model. Source: [38]*

In CNN all layers have a set of neurons and each neuron linked to the subsequent layer neurons to create a network and is known as a fully-connected network. The connections between neurons are represented by weights and the value of it represents its importance. Each neuron calculates the weighted sum of the incoming data and passed through a nonlinear function or activation function and gives an output. That output acts as an input for the next layer neuron and the information transmitted through the network. The last layer where the number of neurons corresponds to the number of classes (in classification task) and predicts the probability of information fit into each class using a nonlinear function. The hidden layers are between the input and output layers. The depth of the model depends on the number of hidden layers but unfortunately, there is no predefined method to select the optimal numbers of hidden layers [38].

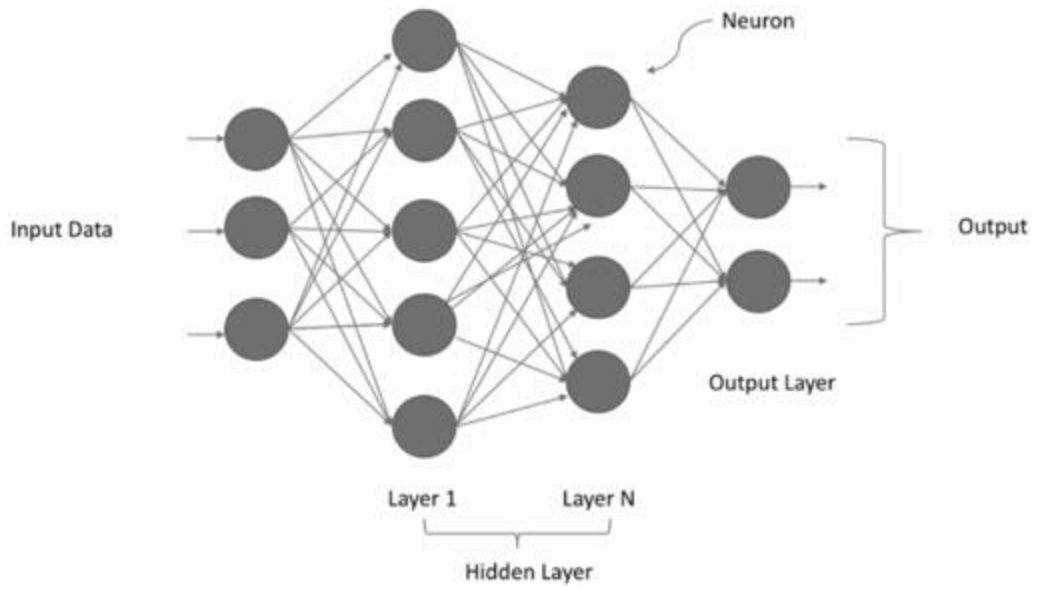

***Figure 19**: Neural Network structure.*

CNN's have the following advantages over the other networks are:

- To extract features from the image, CNN applies direct convolution operation to the pixel of an image.
- CNN's acquire information from a massive amount of data rapidly and represent that information in a distributed manner.
- CNN's have a simple network structure due to sparse interconnection (only a few input units contribute to a given output unit), parameter sharing (same weights are applied to multiple locations in the input), and spatial subsampling that results in more adaptable in image classification.

To better perceive CNN, the following subsection will present the structure in detail [39].

2.5.1 Basic Architecture

The CNN architecture comprises three cascaded layers which are the Convolution layer, Pooling layers, and Fully-connected layers. Convolution block consists of convolutional layers with nonlinear activation functions. The pooling layer is used to downsample the given input to reduce the number of parameters. Both the convolutional and pooling layers are stacked layer by layer as the first few layers in a typical CNN network (see **Figure 20**) and used to gather features from the

input image. Lastly, one or more fully-connected layers are used to gives the class score for each pixel in the input image [40].

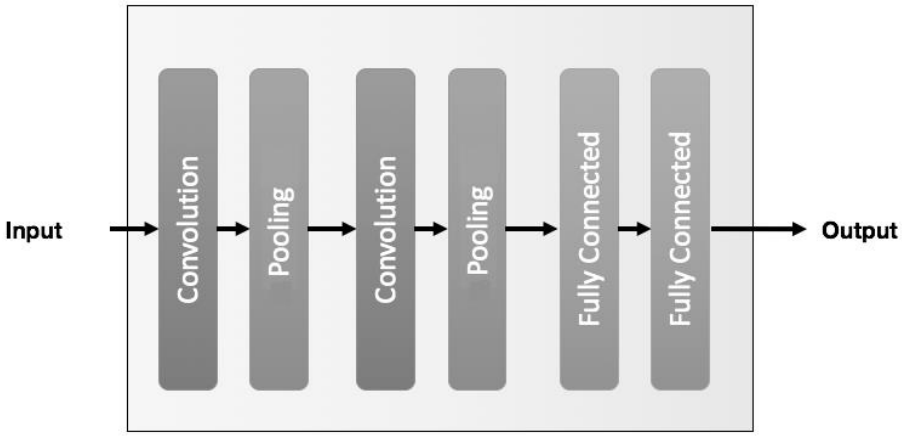

***Figure 20**: Basic CNN Architecture.*

A detailed explanation of each block is given below.

Convolutional Layer

The key element of the CNN network which performs a convolutional operation between the input image and a certain size (usually in squared shape) of the kernel to produced feature map using the activation function. As multiple convolutional layers are stacked layer by layer more sophisticated features are extracted. The kernels usually have a small spatial dimension (width x height) then input feature map but have the same depth. Typically a small kernel size (3×3, 5×5) is used because it significantly reduced the number of learnable and also ensure to learn distinctive patterns from the local region. Kernels are specialized to extract certain features from input image also multiple kernels are applied to a single image to extract multiple features. The number of kernels (usually set between 32 and 512) determines the number of feature maps. By increasing the number of kernels a more strong neural network will develop but also increase the risk of overfitting because of the huge number of parameters. In the convolution operation kernel will positioned over the left upper border pixel of the image and perform dot product between kernel parameters and image pixel and give a single value as a result which shows the presence or absence of a feature in this specific pixel of the image.

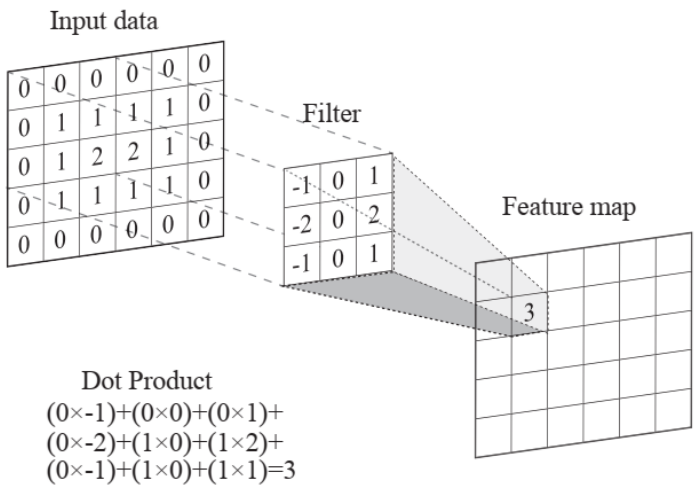

***Figure 21**: Calculation of the dot product of a single-pixel of the feature map. Source: [41]*

After that, the convolution operation slides the kernel to the right and performs convolution operation on the next pixel.

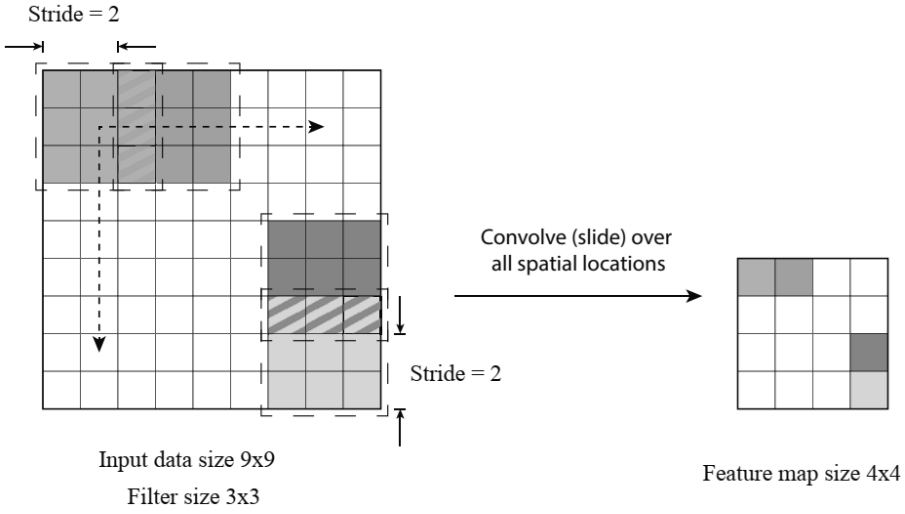

***Figure 22**: The convolution operation. Source: [41]*

The number of pixels kernel window moves is called stride size. The default stride size is 1 which means kernel slide pixel by pixel however stride size can be changed from unity. By increasing the stride size kernel larger steps between pixels which results in a smaller feature map.

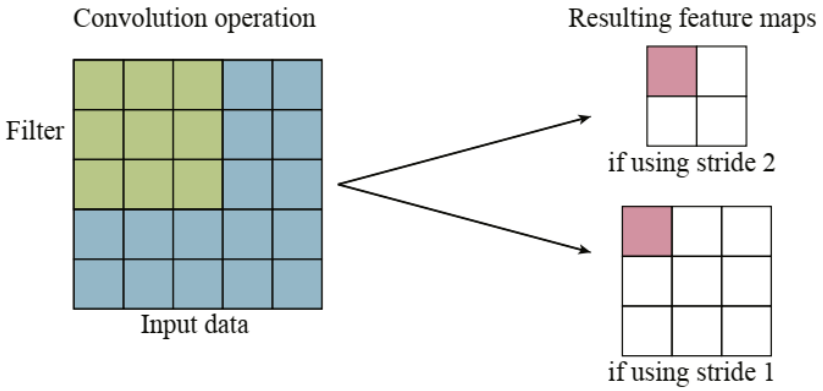

***Figure 23**: Effect on the feature map with different stride size. Source: [41]*

After sliding the kernel on each pixel from the upper left to the bottom right corner, result in a feature map and stored in a spatial grid-like structure.

The convolution operation reduced the output feature map size because kernel central pixel is not positioned at the boundary pixel of the input image and reduces the size of the feature map. Padding is used to retain the original size of the image. In the padding process, zero-value pixels are added around the border of the input image.

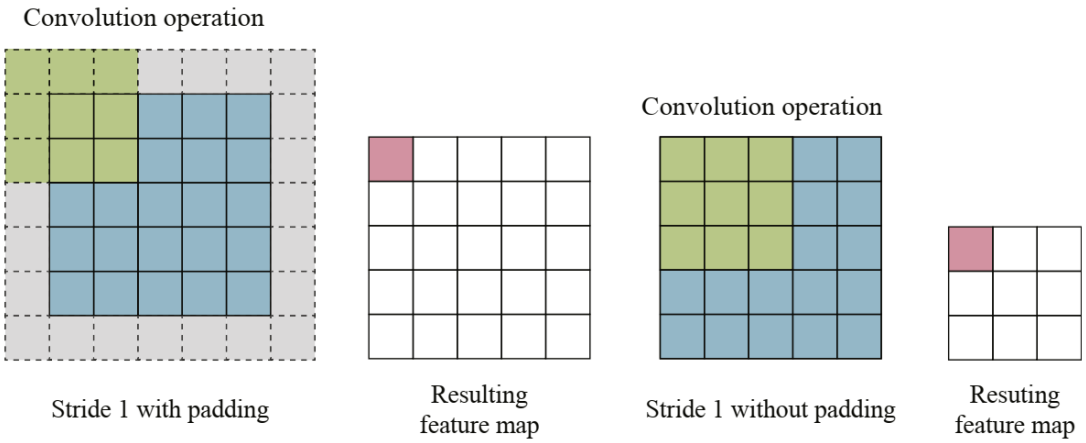

***Figure 24**: The effect of using Padding. Source: [41]*

Two types of padding are valid and same padding. In valid padding, no padding is implied at all and the input image remains in its unaltered shape. In the case of the same padding, a padding layer is added in such a way that the resulting feature map after convolution operation same as the input image [41].

In the neural networks, the main purpose of the activation function is to add non-linearity into the output after the computation of the convolutional operation to learn complex tasks. An activation function is usually defined inside the convolutional layer but sometimes defined as an independent layer for more flexibility in architecture. Commonly used activation functions are the Exponential Linear Unit, Tanh, Sigmoid, and Rectified Linear Unit (ReLU).

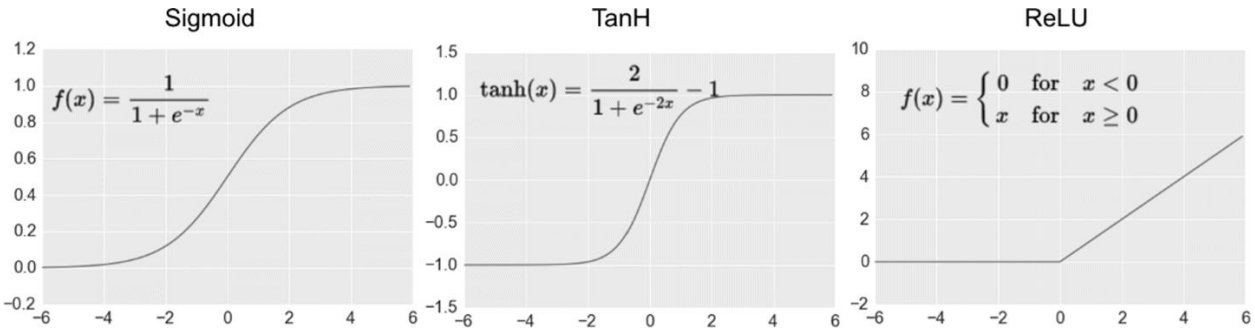

***Figure 25**: Activation Functions, Sigmoid, Tanh, ReLU (from left to right).*

Among all others, ReLU is the most popular because it allows the training of deep CNN much faster [42]. The reason behind that, the ReLU function gradient is non zero for any positive value which helps in faster convergence. But Sigmoid and Tanh function saturates at very high and low values and makes gradient descent close to zero which slow-down optimization process.

Pooling Layer

This layer comes after the convolutional layer and makes another important building block of the CNN architecture. The main objective of this layer is to minimize the spatial dimension of the feature maps captured by the convolutional layer. In addition to that, feature maps counts remain the same as from the previous layer and retain the information to learns important features. As the spatial size of the feature map reduces, the number of parameters and the computational load of the network also reduced. Similar to the convolutional layer, the parameters defined by the user in the pooling layer are the window size (analogous to the kernel size) and the stride size. Usually, in the pooling layer, some pre-defined functions such as maximum, average, and random pooling are used to compile the features.

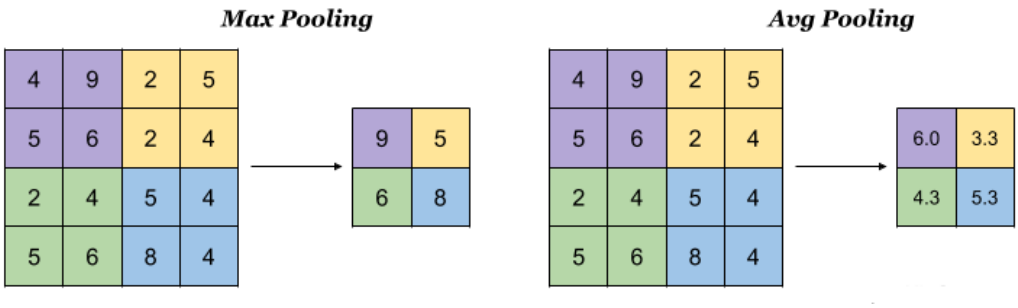

***Figure 26**: Pooling operation with window size 2 and stride 2 (Maximum, Average).*

In maximum and average pooling, pooling layer slides over the input and takes the maximum and average values in the window respectively and all other values in the window are discarded, but in random pooling, values are selected according to a certain probability.

Fully-Connected Layer and Output Layer

The last building block of CNN architecture is the fully-connected layers, usually composed of multiple hidden layers. The feature maps that are extracted are input to the fully-connected layer which flattened these feature maps into a 1D feature vector. Each fully-connected layer contains multiple neurons that are fully interconnected with the following layer neurons. In multiple fully-connected layers configuration, the initial layer takes the input from the last convolutional and pooling layer and apply weights to predict the label.

The weights from the FC layer forward into the classification layer and probability for each class are calculated. For multiple-class problems, Softmax is the most common classifier because it limits the output between 0 and 1 range which is interpreted as a probability, and logistic regression is used for binary classification problems.

In deep CNN usually, a large number of parameters are needed to be estimated to achieve better performance. By increasing model complexity two types of error arise, **Bias** and **Variance**, that's need to be balanced. when the model assumes faulty assumptions due to inefficient or uninformative features, a small number of data samples, and simple algorithm bias arises, and unable to model the relationship between features and true class correctly, this error is known as **underfitting**. On the other side, when the algorithm is complex too many features link to several samples and model gather specific aspects of training data but unable to generalize it, is known as high variance or **overfitting**.

When the model is overfitted it performs perfectly on the training data but not on validation or test data. To minimize the variance error regularization methods are used to prevent overfitting behavior. When the model performs poorly, it is necessary to find whether the model is underfitting or overfitting the data. For an optimized model the error in the training data will be small but in validation set error will be slightly larger than training error.

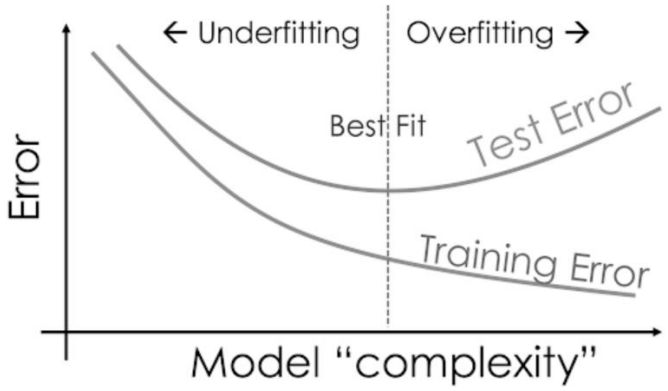

***Figure 27**: Underfit vs Overfit.*

Various regularization techniques are used, for example, L1 and L2 regularization, early stopping, and dropout. In L1 and L2 regularization, weights remain low forcefully (weight values show the importance), now the model is less dependent on training data and generalize better on unseen data [43]. Another simple but effective regularization technique is early stopping the training of the network. Training of the neural network ends when the training set error not decreasing and the validation set error starts increasing, which is a strong indication of overfitting [44]. Dropout is also another powerful strategy in which randomly some of the neurons incoming and outgoing connections are temporarily removed during training. This technique extracts a different set of useful features and increases model generalization ability [45].

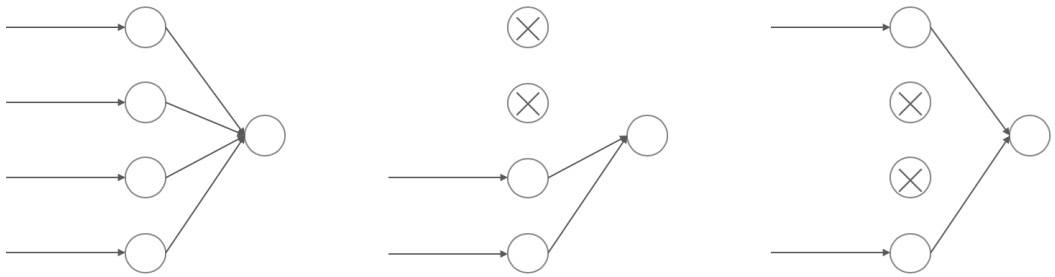

***Figure 28**: Dropout scheme.*

2.5.2 Training and Validation

In supervised learning, training is a process in which the algorithm finds a function that best fits the given input features and target labels or classes. The final goal of the algorithm is to check its performance on unseen data that is not used during training. The training process is divided into three major steps and performs iteratively until the model converged by the adjustment of weights between neurons.

Forward Computation

The first step is forward computation, in which input data is passed through the complete network (convolutional, pooling, and fully-connected layers). Forward computation gives prediction depending upon the network architecture.

Loss Function

The difference between the predicted labels and the true labels is the error, which needs to be minimized by tweaking the network's weights (learned in the forward calculation) and maximize the relationship between input and output. In the second step, the **loss function** is used to quantify these errors in a manner that, smaller the loss function closer the prediction to the true value.

Back Propagation and Optimization

The exact value of the network's weights is estimated using the optimization function called gradient to reduce the value of loss function. The gradient changes the network's weights to a certain extent and measures its effect on output. So, the gradient indicates the direction of changing the weights to minimize the loss function.

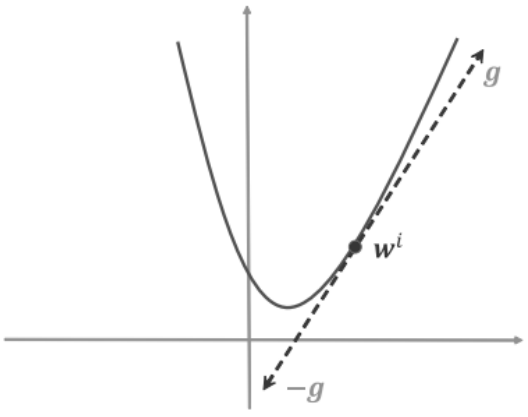

***Figure 29**: Gradient descent method.*

Besides slope, gradient also has a magnitude of the slope which shows how steep the descent is. Now we need to decide the step size we take in the direction, known as the learning rate. It simply defines how fast the network will learn. The learning rate will be chosen carefully because the too small value will cause the model to converge very slowly and large value diverges it [46]. A good practice is, to begin with, a large learning rate and decrease it during the training process.

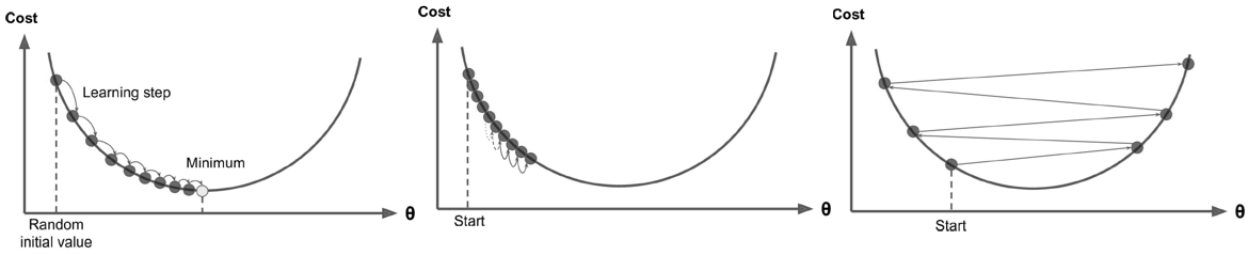

***Figure 30**: Effect of learning rate.*

The magnitude and the learning rate multiplied to obtain final values of weights that need to be altered. Weights are revised and repeat the whole process until the gradient is very small and we reached a minimum. This final step of optimizing weights is known as backpropagation and made it possible to efficiently train the neural network [47, 48].

Several optimization functions are used such as Root Mean Square Propagation (RMSProp), Adaptive Gradient (AdaGrad), and Stochastic Gradient Descent with momentum (SGD) but the most commonly used among all is gradient descent based optimizer is known as Adaptive Moment Estimation (Adam) [49]. Adam calculates the adaptive learning rate for each parameter by using SGD and RMSProp as it is an extension of both.

Parameters and Hyperparameters

The neural network performance depends on the parameters that the model update during the backpropagation step. The core parameters of the neural networks are weights and biases which are updated during training. The parameters which determine the properties of the network and should be set before the training process are called hyperparameters (e.g number of layers, units, learning rate, epoch). Hyperparameter values can not be learned from the training process but tune for a specific problem because hyperparameters are best for one data only and will not performs well on all datasets. To achieve the best performance of the model a range of values of hyperparameters are tested each time and select the one with the best results. The validation dataset is used to find the best hyperparameter values for the model and is referred to as hyperparameter

tuning or optimization. Optimization can be done manually or by an automated process. In an automated tuning, a range of values are tested and select the best out of it but can be computationally heavy. Manual optimization is the most common type and has advantages over the automated tuning of being less computationally expensive but requires technical expertise [50]. Hyperparameters are divided into two types, the first one determines the structure of the network including kernel size, padding, strides, and activation function, the second type determines the training process includes batch size, learning rate, regularization, epochs counts, etc.

3. LITERATURE REVIEW

The following chapter furnishes a thorough review of the existing approaches including deep learning techniques for crop classification especially rice crops using remote sensing. The chapter is divided into three sections. The first section explains the traditional approaches, the second section deals with the conventional methods and finally the deep learning approach using ground truth data or by using spectral unmixing technique before classification in the last section

3.1 Corp Classification Approaches

3.1.1 Traditional Methods

In the traditional approaches, vegetation indices are derived from multitemporal images to monitor crop growth stages to classify and map them. Multitemporal information gives higher accuracy than mono temporal because it gives information on the complete growing season.

M. Usman Liaqat *et al.,* in 2017 reported a comparative study among various vegetation indices e.g. SAVI, MSAVI, NDVI, and EVI in Indus Basin for the wheat yield estimation. A semi-supervised linear regression model was used using remotely sensed data of MODIS and Landsat. In the case of Landsat 8, yield estimated by SAVI proved nice results with R^2 (0.743) and Pearson correlation r (0.8), compared to MODIS. These results suggest that Landsat 8 data with SAVI is the most appropriate and efficient for wheat area mapping and yield estimation under Pakistani's farm conditions [51].

HU Qiong *et al.,* in the same year study the significance of spectral and temporal features for crop classification in Heilongjiang Province China using MODerate-resolution Imaging Spectroradiometer (MODIS) surface reflectance product (MOD09A1) time-series data. Important spectro-temporal features for crop classification are selected using the Separability index (SI). Different vegetation indices (GVI, NDTI, LSWI, EVI, and NDSI) with spectral and temporal features using Support Vector Machine (SVM).

The SI feature selection technique achieves good accuracy of 94.03%. Also, vegetation indices NDTI, LSWI, and EVI provide information about the best temporal window (late August to early September) to identify crops in Heilongjiang. It is concluded from the study that spectral and time

variety is essential for the classification of the crop, and improves accuracy about 30% as compared to the single image [52].

Emma Madigan *et al.,* in 2018 conducted a study using Sentinel-2A/B spectral time series data for two years (2015/2016 – 2016/2017) rice-growing season in the Riverina region of Coleambally, New South Wales, Australia. In that region, 11 rice varieties are grown from which three varieties were used to compare the growth patterns and varietal information using the NDVI and LSWI vegetation indices.

Also using these indices author compares three seed sowing methods (aerial, dry broadcast, and drill) in the Coleambally region because each method adopts a different water management approach. Then he studied their effect on yield in the later growing season which is valuable information and it also enables them to suggest farmers the optimal window for planting/harvesting [10].

Ahsanullah and Sumaira Zafar *et al.,* in 2018, studied the temporal behaviors of different winter and summer crops including their health using vegetation indices. In this study, Sentinel-2 images were taken from October 2017 to April 2018 to cover winter crops (wheat, mustard, carrot, and alfalfa) of Bannu district, Pakistan. Crop differentiation was done by deriving phenological profiles of the crops based on the NDVI and crop calendar. With these phenological profiles of crops and environmental conditions, overall accuracy and Kappa is around 90%. They concluded from this study that mapping of crops by remote sensing data and phenological patterns is an efficient, and cost-effective, technique compared to other methods [53].

Awais Karamat *et al.,* in 2019 predict the rice yield/area for the year 2019 with 10 years (2008-2018) historical datasets of rice crop from Hafizabad, Pakistan. To get more accuracy Crop Reporting Service (CRS) based datasets were integrated with remotely sensed datasets. The Normalized Difference Vegetation Index (NDVI) and Ratio Vegetation Index (RVI) values of each year are computed from spectral responses of rice crops. A linear regression model was applied to NDVI/RVI and the CRS based yield to estimate the rice yield for the year 2019 [54].

3.1.2 Conventional Machine Learning Methods

Conventional or classical machine learning algorithms, for example, Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbour (KNN), and Decision Tree (DT) also used

successfully in crop classification and gives good results as compared to traditional approaches. But a significant effort required to classify crops in terms of domain knowledge and features engineering to extracting features from raw data.

Mrinal Singha *et al.,* in 2016 investigated the object-based classification method for mapping rice fields by using spectral and phenological features. MODIS NDVI time series was used to extract phenology to provide the seasonal dynamics of the crops. On the other hand multispectral data acquired by Environmental Satellite (HJ-1A/B) for the year of 2014-2015, provided the spatial variation patterns. Results obtained by classification were assessed by using 100 sample points confusion matrix. With the proposed method an overall accuracy of 93% was achieved. Furthermore, phenology features refined the overall accuracy from 2%–4%. The underestimated value indifference of 34.53% was noticed in the comparison between estimated rice areas and official statistics. This study highlight that, the combined use of multispectral satellite data and phenology are capable of accurate rice classification in a large area [55].

Markus Immitzer *et al.,* in the same year first time studied the performance of Sentinel-2 data for supervised pixel and object-based classifications during August 2015 in Austria and Germany. Random Forest classifier (RF) was used to classify crop type and tree species. Sentinel-2 data were produced using 10 spectral bands and resampled at 10 m spatial resolution for crop and tree species maps. The overall accuracies were 65% and 76% for tree species and crops respectively. The results achieved for pixel and object-based methods were satisfactory because a single image was available and the full capacity of Sentinel-2 imagery could not be evaluated. The results can be refined with the use of twin Sentinel-2 satellites in the orbit, with increased revisit frequency which permits to exploit novel spectral and temporal information [6].

Y. Guo *et al.,* in 2017 proposed a training method of the sequential classifier on Sentinel-2 multispectral data for mapping rice in New South Wales, Australia, during the 2015-2016 season. The initial flooding period was used to train the classifier and separates rice/non-rice pixels and a trained classifier was then used to aid in the training of classifiers for later rice growing stages. The proposed method was found to be more stable and consistent in rice mapping as traditional spectral signature-based algorithms with accuracies that were greater than 80%. Furthermore, it proved to be computationally efficient and less expensive as a pre-trained classifier used for subsequent computation [56].

R. Saini and S.K. Ghosh *et al.,* in 2018 examined the crop classification using Sentinel-2 imagery in the Uttarakhand, India by using Random Forest (RF) and Support Vector Machine (SVM). Four bands, i.e., Red, Green, Blue, and Near-Infrared of Sentinel-2 satellite were used as input to both classifiers. The overall accuracies are 84.22% and 81.85% achieved by RF and SVM respectively. It has been found that both classifiers performed well, but RF achieved 2.37% more improvement over SVM. This study has shown that Sentinel-2 imagery has the capability for classification and more precise results obtained by RF classifier [57].

Ahmad Khan *et al.,* in 2018 integrating RapidEye and Landsat imagery for mapping and acreage area of winter wheat during the growing season 2014-2015 in Punjab, Pakistan. Four wheat maps (percent cover and categorical training data) derived from Landsat 30-meter spatial resolution are closely matched to official statistics. Based on good accuracies, these maps help in describing the spatial coverage of wheat and have the potential to reduce the standard error of area estimation. This approach is valuable as a comparatively less number of resources are required to implement the method and give pre-harvest crop area estimates [58].

Siwen Feng *et al.,* in 2019 conducted a study in Yushu, Jilin Province, China, to compare the Random Forests (RF) and Support Vector Machine (SVM) methods with a statistical-based maximum likelihood method using Sentinel-2 images during 2017. A total of 126 features (12 spectral bands, 96 texture features, 7 indices, and 11 phenological parameters) are obtained from Sentinel-2A images. Classical machine learning methods (RF, and SVM), and statistical maximum likelihood methods are compared using the best combination of 13 features out of 126 features and obtained overall accuracies of 98% and 88.96% respectively. The short-wave infrared band differentiates rice, corn, and soybean while the water vapor band classify corn and rice significantly. Therefore, classical machine learning methods robustly identify with an accuracy of more than 95% for each crop, but the statistical method of classification results shows imbalanced accuracies for each crop [59].

Licheng Zhao *et al.,* in 2019 put forward a study on the combination of Sentinel-2A images and Unmanned Aerial Vehicle (UAV) images to get finer maps for crop classification in Heilongjiang Province, Northeast China. This unification of yields superior classification results as well as enhance the accuracy of the classification at the scale of a plot. It has been found that on comparing the different spatial resolution of fused imagery, the high spatial resolution doesn't

provide accuracy as noticed in this study 0.10 m resolution UAV fused images generate good results [60].

3.1.3 Deep Learning Algorithms

In the past few years, the use of deep learning algorithms is significantly increased due to its performance in image classification. In this section, deep learning algorithms used for crop classification with ground truth map in the first section and with an abundance map are discussed in the later section.

Deep Learning Algorithms with Ground Truth Map

Nataliia Kussul *et al*., in 2017 studied the multilevel deep learning approach for land cover and crop types classification using multisource multitemporal satellite imagery of Landsat-8 and Sentinel-1A satellites in Ukraine territory. The unsupervised neural network is used to process optical segmentation and missing data rebuilding, and then progress the study based on supervised CNN in which he proposed 1D and 2D CNNs architectures, to discover spectral and spatial features, respectively. In 1D CNN provides a classification of each pixel whereas the 2D CNN performs classification on a patch of pixels of 7 × 7 of the image. Ensembles of 1-D and 2-D CNN outperforms random forest (RF) classifier and ensembles of multilayer perceptron (MLPs) with 94.6% accuracy provide better discrimination in crop types. The main advantage of using CNNs over MLP and RF is that it permits to build a hierarchy of local and sparse [61].

Zhuang Zhou *et al*., in 2018 proposed a CNN based crop classification method and compared it with the Support Vector Machine (SVM) using Sentinel-2A multispectral images in Yuanyang county, Henan province, China. The crops used for classifications are corn, rice, and peanut. Two images of Sentinel-2A during the 2017 season were extracted and input to the CNN model as an image vector and achieve an overall accuracy of 95.6%. For comparison SVM method based on Radial Basis Function Kernel was also used to obtain crop classification results and achieve a good overall accuracy of 94.94% [62].

Yiqing Guo *et al*., in 2018 constructed a deep CNN architecture to classify the rice varieties in New South Wales, Australia using spectral reflectance and phenology. Five rice varieties during the 2016–17 growing season were under study. A time series of optical images (Top of Atmosphere) captured by Sentinel-2A/B satellites on-board Multispectral Instrument (MSI) were

used as inputs of the growing season. A deep convolutional neural network achieved an overall accuracy of 92.87%. Each pixel is used as a single sample and formulated as spectral bands into the imaging dates matrix to improve the classification results. The results achieved were also compared with the traditional approach of Support Vector Machine (SVM) coupled with a Gaussian radial basis function kernel by taking time-series of NDVI and LSWI and found an overall accuracy of 57.49% which was much lower than that of proposed deep CNN classifier [63].

Shunping Ji *et al.,* in 2018 introduced multi-temporal 3D-CNN and compared with 2D CNN and Support Vector Machine (SVM) in pixel-wise classification. Results showed that 3D-CNN is superior for feature extraction in the Spatio-temporal domain. The 3D CNN showed surpass overall accuracy of about 3% as compared to 2D CNN and SVM, and 6% with GF2 data, to distinguish corn from rice. It inferred that the 3D CNN with fine-tuned parameters could not only helpful in crop classification but also for crop growth dynamics and outperformed the current methods. Besides, this study also put forward an active learning strategy to improve the manual labeling accuracy up to a required limit [64].

Meng Zhang *et al.,* in 2018 employed fused multitemporal Landsat 8 and MODerate resolution Imaging Spectroradiometer (MODIS) data with spatial-temporal adaptive reflectance fusion model (STARFM) and use Patch-based deep learning CNN algorithm to signal out large-area paddy rice in the Dongting Lake of China. In the proposed module, time series of the different feature combinations such as Normalized difference vegetation index (NDVI) data, vegetation phenology, land-surface temperature (LST) was used to improve the accuracy of classification. The OA is 97.06% and the Kappa score is 0.91, which is higher than SVM and random forest (RF) method, results also correlated (R2 = 0.9945) with official statistical data [65].

Ziheng Sun *et al.,* in 2020 proposed a deep neural network (DNN) for generating maps during crop season using Landsat-8 imagery for North Dakota. Geoweaver was used to chain all processing workflows in an automated manner. The overall accuracy achieved was above 82%, precision and recall were also very good than Cropland Data Layer (CDL). The proposed model proves to be reusable for bigger regions and generate better in-season maps for major crops as compared to minority crops [66].

Weichun Zhang et al, in 2020 conduct experiments in Banan District and Zhongxian County Southwest hill area in China using Sentinel-2 multitemporal data for mapping rice paddy

area. Four different CNN algorithms (Conv-1D, Conv-2D, two Conv-3D) were developed and also compare with Random Forest, Support Vector Machine, Extreme Gradient Boosting, and Multilayer Perceptron. For each classifier, two schemes, local classification, and a pre-trained classifier were designed. Out of all classifiers and schemes, Conv-2D in local classification experiments performed well in Banan District and Zhongxian County with overall accuracies and F1 scores of 93.14%, 0.8552, and 92.53%, 0.8399 respectively. Among the classical machine learning algorithms, XGBoost performs best in Banan District with an overall accuracy of 89.73%, and the F1 score is 0.7742 and SVM performs best in Zhongxian County with an overall accuracy of 88.57%, and F1 score is 0.7538. Besides that RF and XGBoost attained acceptable transferability while all the CNN models have poor transferability in both experiment schemes [67].

Deep Learning Algorithms with Abundance Map

Alberto Villa *et al.,* in 2011 took the attention to overcome the drawbacks of mixed pixels for the classification of HSI and improve the spatial resolution of images. In this method, soft classification and spectral unmixing algorithms were utilized, to demonstrate the fractional abundances of each class at a sub-pixel level. The spatial resolution has been increased by simulated annealing to locate landcover classes in each pixel. Experiments were done on three different synthetic real data sets and excellent results obtained both qualitatively and quantitatively. The proposed method achieves an overall accuracy of 90.84% which is higher than SVM with an overall accuracy of 88.21% [68].

Inmaculada D´opido and Antonio Plaza in 2011 extract feature using spectral unmixing before going to the supervised classification of urban hyperspectral data. The proposed technique does not penalize classes that are not related in terms of variance or signal-to-noise ratio (SNR) as with principal component analysis (PCA) or the minimum noise fraction (MNF). The results obtained by a reflective optics spectrographic imaging system (ROSIS) over the city of Pavia in Italy, showed that pixel unmixing based strategy provides a suitable alternate to PCA or MNF [69].

Eleftheria A. Mylona, *et al.,* in 2017 introduced an unsupervised spectral unmixing-based clustering method (SUBC) consists of two phases namely: unmixing and clustering with the utilization of abundance representation of high-spatial-resolution hyperspectral images (*HSIs*) at clustering stage. In the unmixing phase, endmember extraction and abundance estimation (AE)

was done and a clustering stage in which pixels with abundance vector were grouped into clusters. Experiments were carried out on a synthetic *HSI* as well as on airborne *HSI* datasets of high-spatial-resolution containing vegetation and urban areas. The results confirmed the effectiveness of the proposed method and revealed its outperformance compared against clustering techniques (k-means, HAC, FCM, and APCM) in terms of overall accuracy (96.3%) and kappa coefficient (0.83) [70].

Eleftheria Mylona *et al.,* in 2018 proposed the methodology for supervised and unsupervised classification of multispectral Sentinel-2 images of Northern Pindos National Park (Greece), with three classes (water, vegetation, and bare soil), utilizing the abundance representation of the pixels of interest. It was obvious from the study that algorithms utilizing abundance representations can precisely evaluate the abundance vectors using a sparsity-promoting unmixing scheme that produces the relevant abundance maps and give more accurate classification maps based on the available reference map [71].

Shuyuan Xu *et al.,* in 2019 develop a methodology by integrating spectral unmixing and classification techniques to characterize the trend of mixed pixels with a physical meaning. This method is based on the premise that the spectral signature of a land cover class is correlated with a given set of endmembers, which define a low-dimensional subspace with a clear physical meaning. It was concluded from synthetic datasets and real HSI that this method performed competitively on limited training sets in comparison with other approaches [72].

Bei Fang *et al.*, in 2020 proposed a design for Hyperspectral Image (HIS) classification jointly using adaptive spectral unmixing and a mixed 3D/2D network with an early-stoping strategy. It would facilitate the classification, which can lessen the computational load and improve the average accuracy of the final classification. Results on four HIS Datasets (Indian Pines, Salinas Valley, KSC, and Pavia University datasets) expressed that the proposed method can perform better than other deep learning and traditional classification methods [73].

4. STUDY AREA AND DATASET

In this chapter, the first part gives the introduction of the study area and the motivation behind the selection of that area. The second part gives details about how ground surveys were conducted in the study area and ground truth data were collected and preprocessed. Also, give details about the acquisition of satellite images, preprocessing, and dataset preparation steps involved using software tools before input to the classification algorithm.

4.1 Study Area

The selected study area is Multan district, Punjab province, Pakistan, comprises of planes in southern Punjab with an area of 3721 km^2 stretching from 29°32'51.4"N 71°00'29.2"E to 30°14'33.2"N 71°49'56.2"E. Punjab has a tropical wet and dry climate, average annual rainfall is around 400mm of which two-third in the monsoon season (June to September) and temperature in the summer season (April to September) vary between 21ºC and 49 ºC and 4ºC and 24ºC during the winter season (December to February).

***Figure 31**: Study area.*

Due to the climate pattern, two cropping seasons in Pakistan are Kharif and Rabi. Kharif is the first sowing season starting from late April to June and ends from October to early December and the main crops are rice, sugarcane, cotton, and maize. Rabi is the second sowing season that begins from late October to December and harvest from April to early May and wheat, barley, mustard, and rapeseed are the main crops. Rice is a water-intensive and tropical plant that requires high temperatures and an excessive amount of water during the growing season. Therefore, Punjab planes and climate suitable for sowing from May to July and harvested from September to November which fulfills temperature and water requirements during summer and monsoon seasons[11]. Punjab province is the most populous and also most productive agricultural land due to the world's largest irrigation infrastructure and contributes about 76% of food production in the country. Rice contributes 0.6% in GDP and about 3.1% of value-added in agriculture. The major rice cultivated areas in Pakistan are Punjab and Sindh province, both produced about 90% of the total country production. In Punjab, the main rice variety is Basmati rice, other rice varieties like non-Basmati (IRRI-6/9) and hybrid are also cultivated. But in Sindh province, the main variety used to cultivate is IRRI-6 and IRRI-9. Besides, some other varieties are also cultivated. In the Multan district, both major varieties of rice Basmati and IRRI are cultivated due to geographical area and climate [12].

4.2 Dataset

4.2.1 Ground Truth Data

During the rice-growing season of 2019 ground surveys were conducted in the steady area and rice variety, and field geospatial location data was collected using Geographical Information System (GIS) application. A total of 12 rice fields with an area of more than 307 acres are selected for surveys at a different geographical location. Two rice fields areas are shown in **Figure 32**.

***Figure 32**: Ground survey fields.*

All the rice fields polygons are reprojected into UTM/WGS84 zone 42N Coordinate Reference System (CRS) encoding from the European Petroleum Survey Group (EPSG)-4236 into EPSG-32642 using QGIS application for further processing. Pakistan falls under the northern zone from 41N to 43N and our study area belongs to 42N zone. UTM zones must be select or converted carefully, distortion, ill-mapping, and shift results if UTM zones are not selected properly.

4.2.2 Satellite Imagery Data

Sentinel-2 Level-2A Product

In this study, the Sentinel-2 level 2A product was used because of Copernicus open data policy, good spatial, spectral, and temporal resolutions as compared to Landsat and MODIS. In the study area, the rice crop cycle starts from May to July and harvested from September to November. Therefore, the level-2A product during that period was downloaded for the year 2019. Total 53 products were selected after excluding the products with a cloud cover of the study area.

Open Access Hub or SciHub

All the products are freely available at Copernicus Open Access Hub in Standard Archive Format for Europe (SAFE) files. Open Access Hub is a map browser for downloading Sentinel mission data through the graphical user interface [74]. It allows the user to search for data by sensing time, ingestion time, area, sensor type, product type, sensor mission, sensor mode, orbit number, and cloud cover. The user can draw rectangle or polygons of the area of interest (AOI) using a drawing

tool. All the images belong to the area of interest are shown in the map window and list on the left side in the result window with download URL in ZIP format as shown in **Figure 33**.

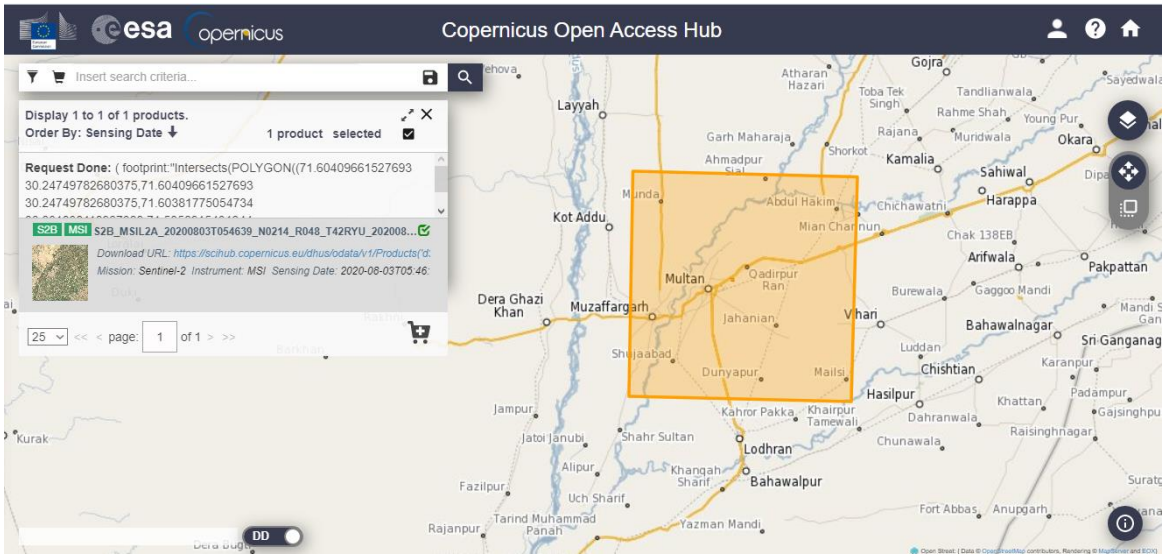

***Figure 33**: Open access hub browser window.*

The SAFE format containing image data in binary and metadata in XML format. The product group together several types of files which are:

- Metadata files (XML file)
- Image preview (JPEG2000 with geolocations)
- Each tile files with reflectance image data files in BOA (Geographic Markup Language (GML)/JPEG2000)
- Datastrip files
- Auxiliary data (International Earth Rotation & Reference System or IERS, Digital Elevation Model or DEM)
- Ancillary data (Ground Image Processing Parameters or GIPPs)

4.3 Data Pre-Processing

All the Sentinel-2 data pre-processing steps were carried out using the Sentinel Application Platform (SNAP) which is concertedly developed by Brockmann Consult, SkyWatch, and C-S. It is a free distributed public tool for satellite data processing. SNAP due to its modular rich client platform, portability, tiled memory management, and graph processing framework makes it an

ideal platform for earth observation data processing and analysis tools. The interface of ESA SNAP with workspace, navigation pan, and RGB image window is shown in **Figure 34**.

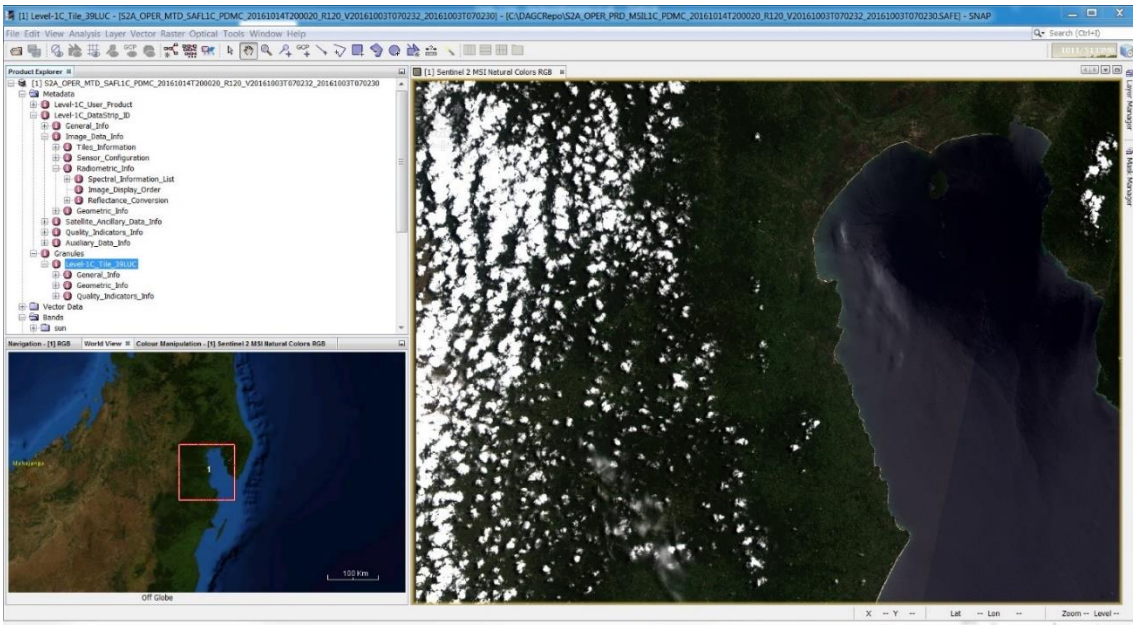

***Figure 34**: Main window of ESA SNAP software.*

SNAP interface can be configured from View panel in the main toolbar by the user according to preference.

4.3.1 Resampling

After downloading all the products the next step is the resampling because Sentinel-2 spatial resolution varies between 10m to 60m depending on the band as shown in **Table 3**. In SNAP software most of the functions require uniform spatial resolution for every band in the source image. Therefore all the bands with spatial resolution greater than 10m were upsampled from coarser to fine resolution using the Nearest-Neighbour interpolation method. The nearest neighbor method is used where data is in a discrete pattern or has distinct boundaries. In that method, each pixel value is calculated using the nearest pixel values in the source image. Bilinear and Cubic upsampling methods are also available in the SNAP.

4.3.2 Reprojection

Transforming the data from one coordinate reference system to another is the reprojection process. Sometimes it is unnecessary if the data is already in the correct coordinate reference system and

used only when defined incorrectly or unknown. Sentinel-2 data by default are in UTM/WGS84 coordinate reference system. As the downloaded products are already in the correct coordinate reference system so no reprojection was applied to the Sentinel-2 imagery.

4.3.3 Masking Area Of Interest

Each Sentinel-2 image covers an area of 100x100 km^2 which requires longer processing time. Therefore first clipping or subset applied to limit the spatial range to our study area which reduces the area and file size. Also, only the necessary bands are selected which reduces the processing time significantly. Secondly, the mask is applied to the subset data using rice fields polygon or vector shapefiles created in the ground survey data. In the masking process Shuttle Radar Topography Mission (SRTM) digital elevation model was used with a spatial resolution of 3 arc-second (90 meters). SRTM model is used to fill the missing area in the elevation model and correct the topography of the area.

4.3.4 Vegetation and Water Indices

Vegetation indices are very important to indicate a feature of interest using two or more spectral bands. In that work three most commonly used vegetation and water indices are used for classification and mapping of rice varieties. These indices are Normalized Vegetation Difference Index (NDVI), Normalized Difference Water Index (NDWI), and Normalized Difference Moister Index (NDMI). All these indices are calculated using build-in vegetation and water indices processor in SNAP, also these can be calculated using band math by defining math expression (see **Table 1**) for each index. NDVI is used to quantify the vegetation using near-infrared and red bands by measuring the difference between them [28]. These two bands are used because healthy vegetation strongly reflects near-infrared light and strongly absorbed red wavelength light. The range of NDVI varies between -1 to +1, a value close to +1 indicates dense green vegetation, and close to 0 is for deed vegetation. NDWI is used because rice fields are flooded in the transplantation method which is mostly used in Pakistan for rice cultivation. Therefore, NDWI is used which is very useful in detecting and mapping water bodies because water strongly absorbed light from visible to infrared spectrum [30]. NDWI uses green and near-infrared bands (see **Table 1**) to quantify water bodies ranging from 0 to 1. A value greater than 0.5 corresponds to water and below that are for vegetation and buildup areas. The third index which is used is NDMI is used to quantify water content in vegetation canopies using near-infrared and shortwave infrared bands (see **Table**

1) [31]. The shortwave infrared band reflects changes in vegetation water content and spongy mesophyll structure in canopies but the near-infrared band is affected by leaf internal structure. The NDMI help in recognizing the area of the field with water stress problem and values ranging from -1 to +1. Values close to +1 show a full canopy cover with no water stress and close to 0 are less canopy cover and water-stressed. NDMI negative values are used to indicate less or no canopy and wet or bare soil conditions. All the resulting NDVI, NDWI, and NDMI products were merged with the other spectral bands using Band Merged tool in SNAP as GeoTIFF format.

4.3.5 Spectral Unmixing

As discussed in section 5.1, the spectral unmixing technique is used to deal with mixed pixel and give sub-pixel level information which is not available in the ground truth map. As the rice fields and the surrounding areas are in a spatially discrete pattern, the linear spectral mixing model was used to extract pure spectra and contribution of each endmember in mixed pixel. The first step for spectral unmixing is the extraction of the pure endmember and image endmembers method was adopted for extraction in that work because it has the advantage of similar atmospheric conditions as the satellite sensor. Pure endmembers for each class are extracted using the SNAP Spectrum viewer tool by identifying the spectrally homogeneous region of each class. In this work, three classes namely Basmati, IRRI, and Other Land pure endmember are extracted. After extracting pure endmembers abundance map was constructed for each class using the SNAP Spectral Mixing tool. In spectral unmixing extracted pure endmember of each class and source bands with the minimum bandwidth are provided and the linear model calculates the abundance fraction of each class in every pixel. For each class, a separate abundance map was created for further processing in the dataset preparation process. Endmembers extraction and their spectral response pattern are shown in the figure below.

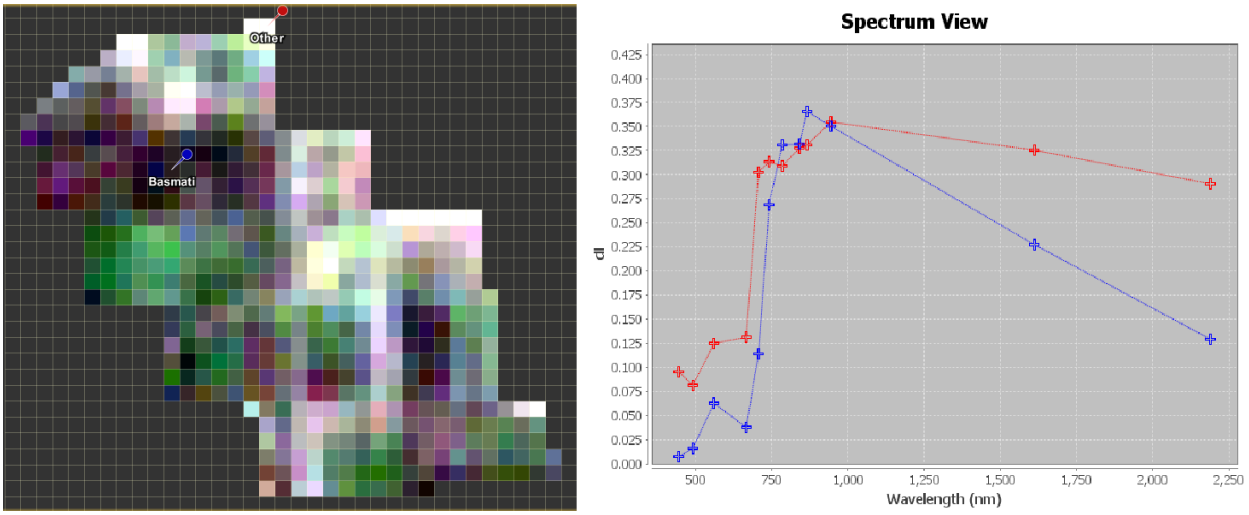

***Figure 35**: Rice filed pure endmember extraction and their spectral signature.*

All the above pre-processing steps working details in SNAP are given in Appendix A.

4.4 Dataset Preparation

In this study, all the datasets were prepared in the Spyder 3 environment using Google open source library Tensorflow and Keras neural network library. all the pre-processing step was completed, masked rice field maps along with vegetation indices, and the abundance map form spectral unmixing for each temporal resolution are available. Dataset images are constructed in such a way that each pixel represents a two-dimensional image with size k × t, (height × width) where k is the spectral and indices bands and t is the temporal resolution as shown in the figure below.

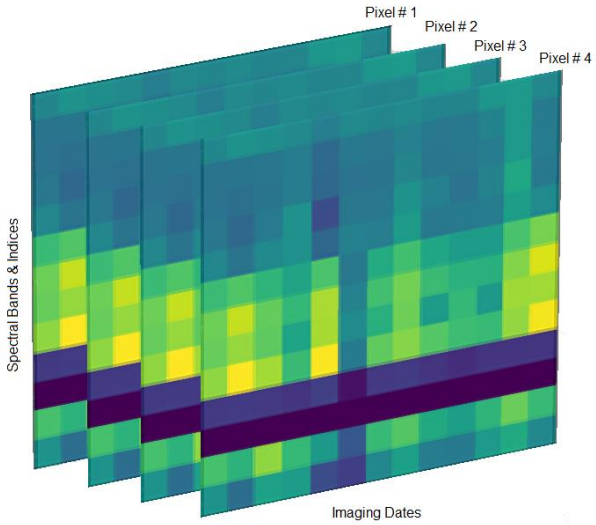

***Figure 36**: Image construction from the pixel's spectral and temporal resolution.*

5. METHODOLOGY

In this chapter, the adopted methodology describes in detail the implementation of the deep learning algorithm CNN. In the first section, the spectral unmixing, it's model, and abundance representation was discussed. In the second section, CNN architecture of the classification model is explained, network training procedure, experimental settings, and in the last performance metrics are discussed to evaluate the performance of the classification model.

5.1 Spectral Unmixing

When dealing with remotely sensed data the most important thing is the pixel size or ground instantaneous field of view (GIFOV), which is the smallest unit and represents a certain area of ground in the image. pixels in the image are arranged in a two-dimensional structure and all are identical in size and shape. Each pixel has a single spectral pattern or radiation intensity level represented by a digital number (DN). Now the problem arises with the low spatial resolution imagery as the size of pixel or IFOV is larger than the object or area of interest and unable to separate the pure signature of each object or feature. Now each pixel spectral signature is a mixture of several materials or features individual pure spectral signature also referred to as endmembers (water, soil, rock, vegetation, etc), and the fraction of each in the pixel is known as fractional abundance [75, 76]. Each mixed pixel is a combination of two or more classes that severely affect the classification accuracy when performing supervised classification [77].

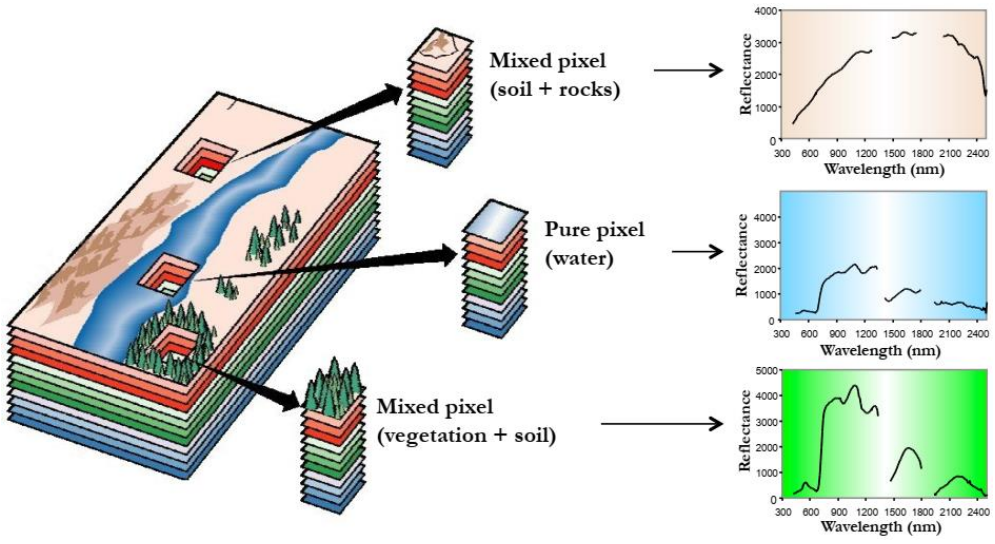

***Figure 37**: Pure and mixed pixels. Source: [78]*

Spectral unmixing is a feasible technique for the analysis of sub-pixel information and indicate each endmember proportion present in the pixel. In the spectral unmixing process, the measured spectrum of a mixed pixel is disintegrated into several distinct spectra or endmembers. After determining each endmember in the image abundance map is constructed for each endmember which indicates the fraction of each endmember present in the pixel. Two models used in spectral unmixing are linear and non-linear mixing. The present study is focussed on the linear mixing model.

5.1.1 Linear Mixing Model

In a supervised classification linear mixing model is used to retrieve endmembers and estimate their abundance without loss of data. The linear mixing model is used in a situation where endmembers are arranged in a discrete, spatially segregated, or checkerboard pattern in a pixel. In this case, the reflected energy will only give the characteristics of the related endmember with the same proportion.

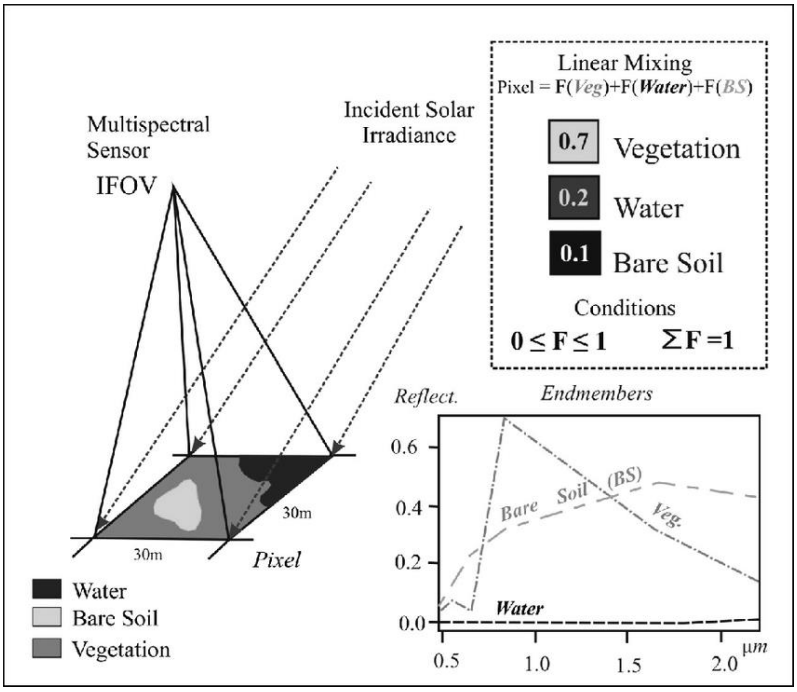

***Figure 38**: Linear mixing model. Source: [79]*

Therefore, a linear relationship exists between the fractional abundance of endmember and the spectra of the reflected radiation, hence known as the linear mixing model (LMM). Mathematically,

$$X = \sum_{i=1}^{M} a_i s_i + w = Sa + w \qquad \textit{Eq. 5.1}$$

Where *X* is the received pixel spectrum vector, *s* is the spectral response of each endmember, *a* is the fractional abundance vector, and *w* is the noise vector. Two physical constrains on the model are non-negativity (all abundance values must be nonnegative) and full additivity (all abundance sum must be equal to unity).

$$a_i \geq 0 \qquad \textit{Eq. 5.2}$$

$$\sum_{i=1}^{M} a_i = 1 \qquad \textit{Eq. 5.3}$$

The first equation represents non-negativity constraint and the second is the sum to unity constraint. The linear mixing model consists of three-steps, dimensionality reduction, endmember extraction, and abundance determination or inversion. Dimensionality reduction is the optional step and used to reduce the computational load of the algorithm. Our focus is on the other two processes [75] [80].

5.1.2 Endmember Extraction

In endmembers extraction first task is to find the pure spectral signature (a pixel composed of 100% single endmember spectral signature). The endmembers can be selected either directly from the image also called image endmembers (IE) or from field or library is known as library endmembers [81]. In image endmember, a pixel is selected from the image with the maximum abundance of the endmember or using algorithms and field knowledge. The advantage of image endmember over library endmembers is that the spectral signatures are obtained under the same conditions as the sensor. Library endmembers are more suitable when estimating abundance across scenes and acquire under different conditions and time of year.

5.1.3 Abundance Determination or Inversion

After the endmembers are extracted, an abundance fraction for each pixel is determined. Different algorithms are used to estimate the abundance vector *a* in the LMM such as least square (unconstrained, fully constrained), and minimum variance method. **Unconstrained least square method** is the easiest one to estimate the abundance vector *a*. In this method, the noise vector w is considered equal to zero and the abundance of each endmember is found after the spectral

response matrix *s* of endmembers is known. The unconstrained method is computationally light because physical constraint (non-negativity, full additivity) does not assume and values lower than zero and greater than unity will appear in abundances. **Fully constraint method** efficiently satisfied both physical constraints and least square error and the most widely used technique for endmember abundance estimation[82].

5.2 Classification

As a deep learning algorithm was selected for this work because it extends the classical machine learning approach by adding convolutional layers and hierarchically represent input data. For the rice crop classification, a 2D CNN was considered, also the architecture we have chosen is based on related work in [64] and will be referred to as **Conv2D** onwards. The followed network architecture was derived from **Oxford's Visual Geometry Group (VGGnet)** for CNN training, but some changes are made to the default architecture to get the best model for rice classification [83]. The following section explains the adopted architecture.

5.2.1 CNN Architecture

All the CNN architectures follow the same architecture, after input, it starts with multiple convolutional and pooling layers, and end with fully-connected layers. The input image is a two-dimensional image with size k × t, (height × width) where k is the spectral and indices bands and t is the temporal resolution. Conv2D first six layers consist of three convolutional and three max-pooling layers. In the first two convolutional layers, the filter size is 32, and in the third layer, filter size changed to 64 to increase learnable parameters in the hidden layers. Also in the first two convolutional layers, the size of the Kernel is 3 × 3, but in the last layer, small size of 1 × 1 is used. In all the three max-pooling layers, the pooling size is 2 × 2, and stride 2 is used to reduce the spatial dimensions of the input data and also the computational load. The rectified nonlinear activation function or ReLU is used because of its performance over other activation functions and introduced non-linearity in each convolutional layer [42].

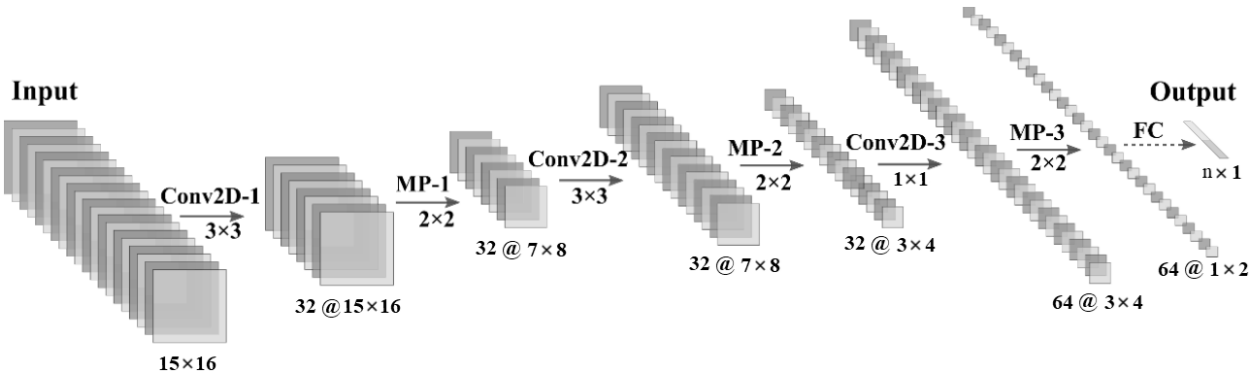

***Figure 39**: Conv2D architecture.*

After that, six fully-connected layers are used with dimensions of n × 1 in the last fully-connected layer, where n is the number of classes. In between two dropout layers are used with a 20 percent dropout as a regularization technique to improve the model generalization ability. The last fully-connected layer uses the Softmax activation function to give the probability of the input image predicted class. The Conv2D model layers summary is given in the table below.

***Table 5**: Conv2D model layers summary.*

Layers	Filter Size	Kernel Size	Feature Map Size	Parameters
Input Image	-	-	15 × 16	-
1st Conv2D	32	3 × 3	15 × 16	320
Max-Pooling	-	-	7 × 8	0
2nd Conv2D	32	3 × 3	7 × 8	9248
Max-Pooling	-	-	3 × 4	0
3rd Conv2D	64	1 × 1	3 × 4	2112
Max-Pooling	-	-	1 × 2	0
Flatten	-	-	128	0
Dropout	-	-	128	0
Dense	-	-	64	8256
Dropout	-	-	64	0
Dense	-	-	32	2080
Dense	-	-	3	99
Total Trainable Parameters				22115

5.2.2 Network Settings and Training

Before starting training each class images arranged into proper directories because Keras high-level library ImageDataGenetartor is used. In ImageDataGenerator flow_from_directory method is used to read the images from containing folders. All the images are categorized into training and validation sets and create the directories, classes folder as shown in the figure below.

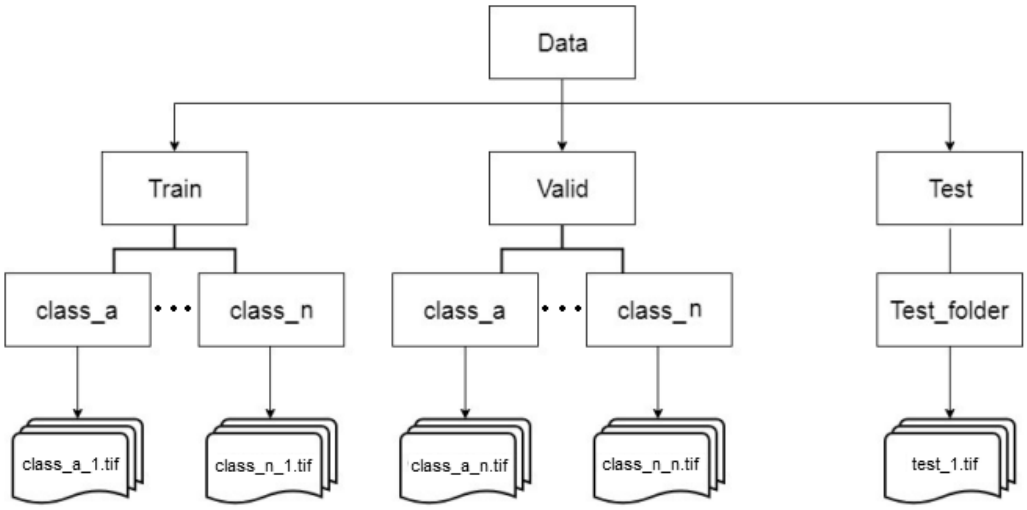

***Figure 40**: The directory structure for classification.*

As shown in the figure train and valid folder contain subfolders with a class name but in the test directory subfolders are unlabeled and leave for classification algorithm to predict the label for each test image.

The procedure of training is shown in **Figure 41**. Pixel-based images with the true labels are input to the CNN classification model and the Softmax activation function is to predict that labels. After that, the cross-entropy loss is computed to compare the predicted and true labels and backpropagated the loss value using Adam optimization function to update the parameters.

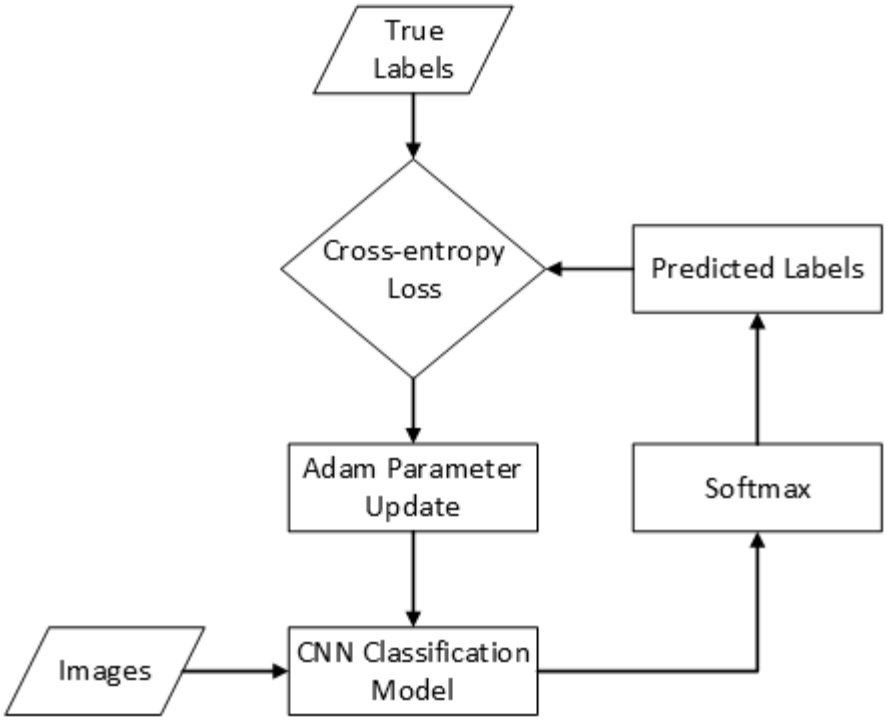

***Figure 41**: Training procedure of the CNN classification.*

In this study, the experiments were built on Google open source library Tensorflow and Keras neural network library in Spyder 3 on a system with Intel(R) Xeon(TM) E5-1650 server processor with 32 GB of RAM and Nvidia Geforce GTX 1660 Ti GPU.

The input data is divided into training and validation set with a distribution ratio of 70 and 30 respectively. Each input image consists of 12 spectral bands, 3 vegetation indices, and 16 temporal dimensions which cover the complete growing season of Basmati and IRRI rice. Total ground survey area pixels are 12442 and each pixel represents an image in spectral and time domain. Test rice fields are selected randomly and consist of a total of 1486 images. Training images were used to train the model and validation images to tune hyperparameters by considering the overall accuracy as the evaluation criteria. In all the experiments the network trained for 100 epochs using Adam optimization function. Additionally, all the images in the training and validation set are shuffled before each experiment.

Hyperparameters

In CNN classification many hyperparameters need to be tunned during the training process to get the most generalized model. The hyperparameters that are tunned during the experiments are batch

size, filter size, kernel size, and dropout percentage. During experiments, only a single hyperparameter value was changed and all others are unchanged to get the optimal value. Hyperparameters variations are shown in the table below.

***Table 6**: Hyperparameters values during experiments.*

Hyperparameter	Value
Batch Size	32 ~ 128
Filter Size	8 ~ 256
Kernel Size	1 × 1 ~ 5 × 5
Dropout	10% ~ 50%

Batch size is the number of training images used to train the network in a single iteration. Batch size is used to estimate the gradient and update the model parameters [84]. In that work, different batch sizes are used for training and validation sets. The second hyperparameter which was tunned is filter size or the number of feature maps, its value varies ranging from 8 to 256. In each convolutional layer, the different filter size is used to increase the feature map and trainable parameters in hidden layers. Small kernel size is used because of small image sizes, ranging from 1 × 1 to 5 × 5, and varies in each convolutional layer. The last hyperparameter is dropout which is also used to avoid overfitting problems by temporarily remove some of the neurons randomly depending upon the percentage of dropout. Different experiments were done using dropout range from 10 to 50 percent. After applying different hyperparameter schemes the best model was adopted which are described in section 5.2.1.

5.2.3 Performance Assessment

After the model is trained and tested the next step is to quantify its performance. Different metrics are used to assess the algorithms because of each metric highlight different aspect of the model. Some of the most prominent evaluation metrics used for classification are introduced here [85].

Confusion Matrix

The confusion matrix is the most intuitive because from this several matrices are calculated to evaluate algorithm performance. In the confusion matrix, a table is used to visualize the performance of the model on test data with known true values. In the table, classes are broken

down into each class and true or false predictions are summarised as counts. The numbers inside the tables are more important because it tells us what type of errors are being made by the model.

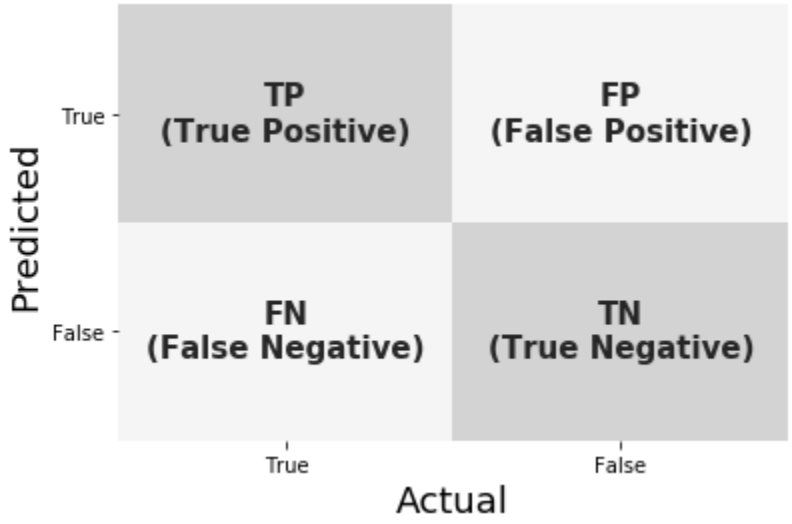

***Figure 42**: Confusion matrix.*

The terms associated with the confusion matrix are:

- **True Positive (TP):** when the actual label belongs to a true class and the predicted label is also true (correctly predicted as a positive class).
- **False Positive (FP):** when the actual label belongs to a false class and the predicted label is true (incorrectly predicted as a positive class).
- **False Negative (FN):** when the actual label belongs to a true class and the predicted label is false (incorrectly predicted as a negative class).
- **True Negative (TN):** when the actual label belongs to a false class and the predicted label is also false (correctly predicted as a negative class).

False positives are also called **Type-I error** and false negative are **Type-II error**.

Accuracy

Accuracy is the most commonly used metric and referred to as the total number of corrected predictions out of all predictions.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \qquad Eq.\ 5.4$$

The numerator has the corrected predictions and all the predictions are in the denominator. The accuracy metric is a good measure only when the data is balanced and does not account for imbalanced classes. For imbalanced data, the more appropriate metric is balanced accuracy which takes the average accuracies for each class.

$$Balanced\ Accuracy = \frac{\frac{TP}{TP + FN} + \frac{TN}{TN + FP}}{2} \qquad Eq.\ 5.5$$

Precision

Another important metric that is also derived from the confusion matrix is precision also referred to as positive predictive value (PPV). It tells us what proportion of predicted positive was actually positive?

$$Precision = \frac{TP}{TP + FP} \qquad Eq.\ 5.6$$

The expression of precision is a ratio of predicted true positive with predicted total positive (true positive and false positive), i.e Type-I error.

Recall or Sensitivity

Recall also called sensitivity gives the true positive rate (TPR) by defining as the ratio correctly predicted positive or true positive and the total actual positive.

$$Recall = \frac{TP}{TP + FN} \qquad Eq.\ 5.7$$

If we want to minimize false negative we focus on recall and ignoring precision while for false positive we focus on precision to make it closer to 100%. Both precision and recall are used to appraise the model performance when the classes are imbalanced.

Specificity

Specificity gives us the true negative rate (TNR) which is the exact opposite of recall. It measures the proportion of the predicted negative that are actually negative values (false positive and true negative).

$$Specificity = \frac{TN}{TN + FP} \qquad Eq.\ 5.8$$

F1 Score

F1-Score also known as F-measure is represented as the harmonic means of precision and recall. It gives a single score which varies between 0 and 1 and represents both precision and recall. F1-score helped us in balancing both precision and recall.

$$F1 = \frac{2TP}{2TP + FP + FN} \qquad \textit{Eq. 5.9}$$

F-measure use harmonic means instead of arithmetic mean because it is more reasonable when computing means of ratios [86].

Cohen's Kappa Coefficient

Cohen's Kappa score is a very useful metric that can handle multi-class classification and imbalanced class problems. Coefficient gives a value less than or equal to 1, less than zero value indicates that the model is useless. It is defined as

$$\kappa = \frac{Po - Pe}{1 - Pe} = 1 - \frac{1 - Po}{1 - Pe} \qquad \textit{Eq. 5.10}$$

Kappa score measures the agreement between two raters, where *Po* is the observed agreement and *Pe* is the expected agreement [87].

6. RESULTS AND DISCUSSION

6.1 Classification Model Performance

After performing different experiments to tune hyperparameters, the best combination was implemented to classify rice varieties during the growing season 2019. Different performance metrics are used to evaluate the classifier namely confusion matrix, overall accuracy, precision, recall or sensitivity, F1 score, and kappa score. Two rice fields are selected for testing purposes, one from each variety of rice, Basmati, and IRRI.

6.1.1 Confusion Matrix

Almost all the metrics are derived from the confusion matrix so we discuss it first. Normalized confusion matrix of the individual Basmati and IRRI class and the combined of all classes are shown in **Figure 43** and **Figure 44** below. From confusion matrix results it is clear that Other Land class mixed with both rice varieties Basmati and IRRI because Other Land is non-cultivated areas with in the rice fields and spectral signatures, indices, and shadow affect the classifier performance.

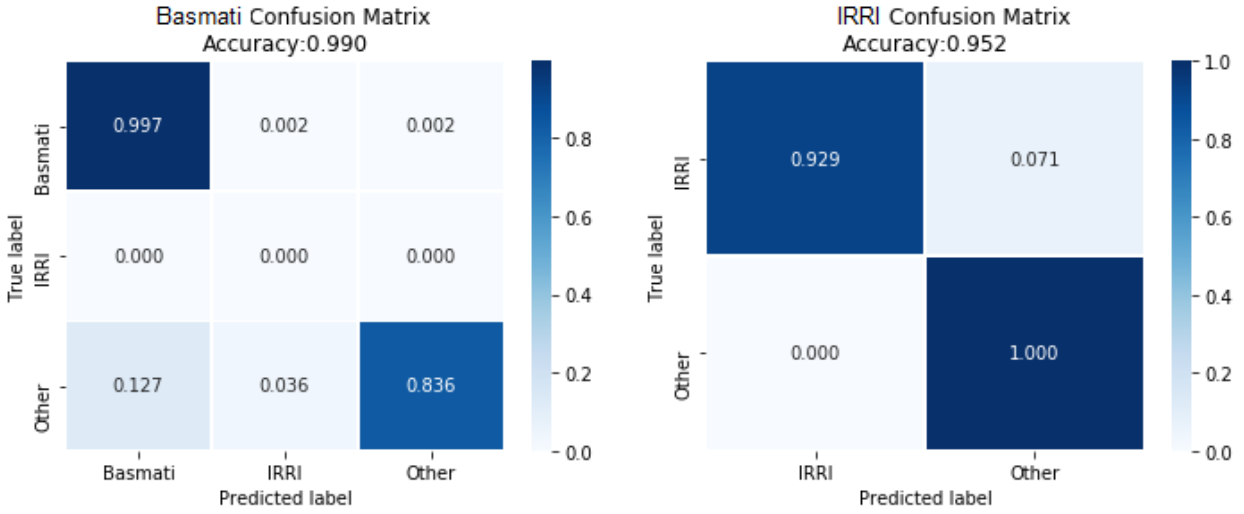

***Figure 43**: Normalized confusion matrices of the individual class of Basmati and IRRI.*

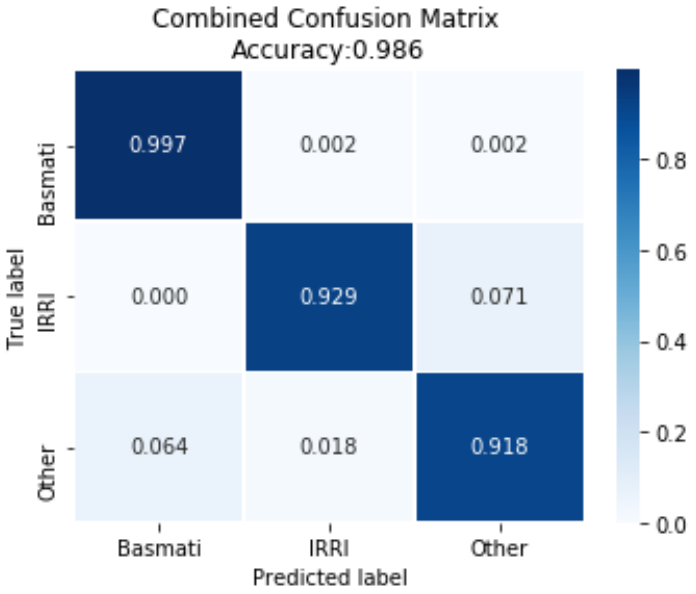

***Figure 44**: Normalized combined confusion matrix of all classes.*

6.1.2 Performance Metrics

The classes in which we are interested are Basmati and IRRI, and the pixel-based CNN classifier achieves an overall accuracy of 98.6% and the same for precision, recall, and F1 score. Also, the Kappa score is close to 1 which is interpreted as nearly perfect agreement.

***Table 7**: Classification results of the individual and combined classes.*

Class / Metrics	Basmati (%)	IRRI (%)	Combined (%)
Accuracy	99.7	95.2	98.6
Precision	99.7	95.2	98.6
Recall	99.7	95.2	98.6
F1 Score	99.7	95.2	98.6
Kappa Score	87.4	89.6	94.6

Performance metrics are also calculated for an individual class and Conv2D shows the best performance in Basmati class with an overall accuracy of 99.7% as compared to IRRI with an overall accuracy of 95.2%. The one reason behind the low performance of Conv2D in IRRI class is its low temporal resolution due to cloud cover in the growing season and Conv2D not able to

learn some missing features as compared to the Basmati class. Another reason is that rice varieties seeds are mixed with other varieties that change the spectral characteristics and performance of the classifier. Kappa score for both classes is above 85% which is also interpreted as excellent agreement.

6.2 Classification Maps

For both rice varieties, Basmati and IRRI ground truth map (generated from abundance representation), the Conv2D classification map, and the classification confidence level map are generated and shown in **Figure 45** and **Figure 46** below. The white pixels in the maps represent No Data values because of masking only the area of interest. Basmati rice is represented by dark green and IRRI with a light green color. The land which is non-cultivated within the rice fields are represented by brown color and named as Other Land.

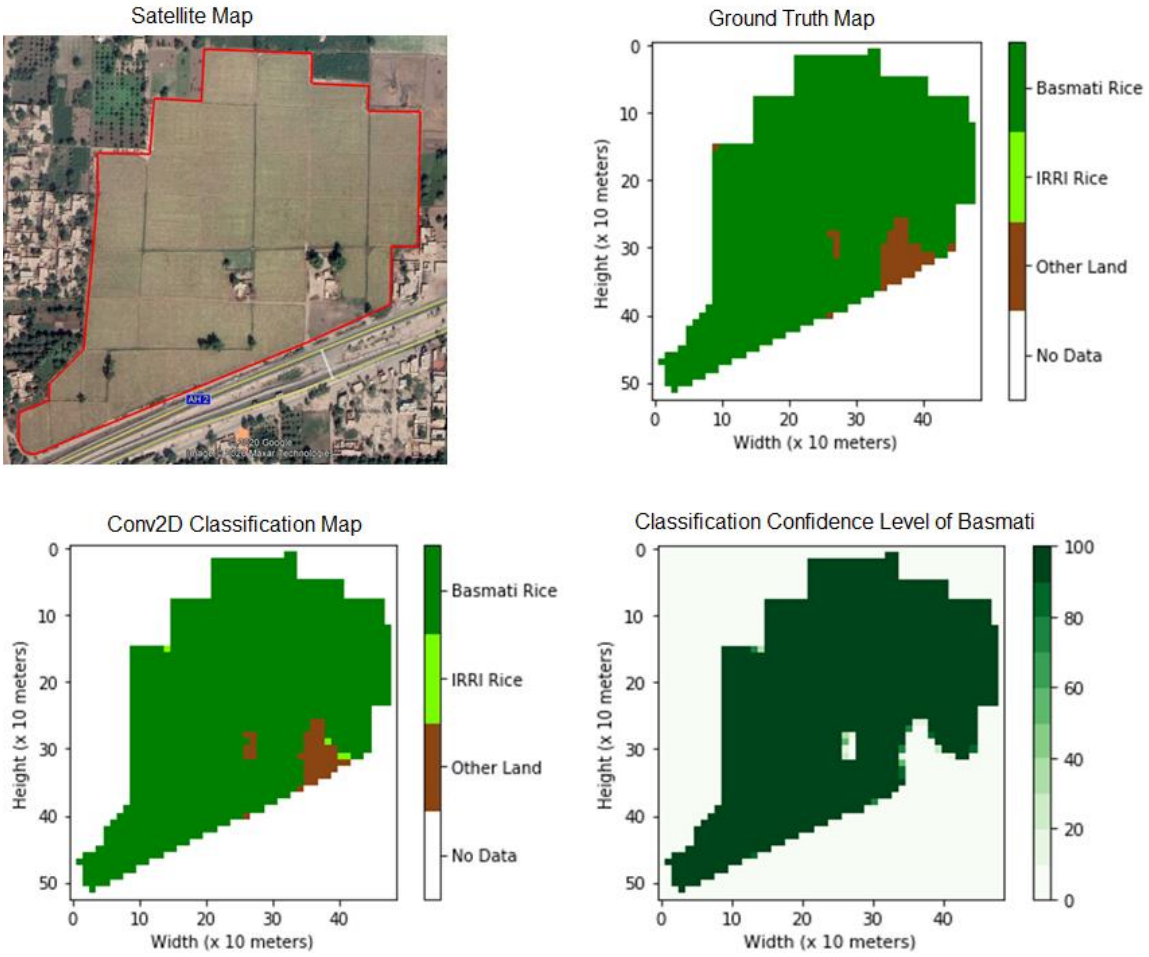

***Figure 45**: Basmati rice satellite, ground truth, classification, and confidence level map.*

In **Figure 45**, the Basmati Conv2D classification map shows some pixels are miss classified as Other Land or IRRI rice type, the same as in **Figure 46** some pixels are miss classified as Other Land but not a single pixel in Basmati class. The reason behind that some farmers grow double crops of rice and the growing season of IRRI rice is after winter wheat but Basmati season starts during or after IRRI. Therefore, in some areas of Basmati fields, IRRI rice also grown with it and mixed with spectral signatures of Basmati rice.

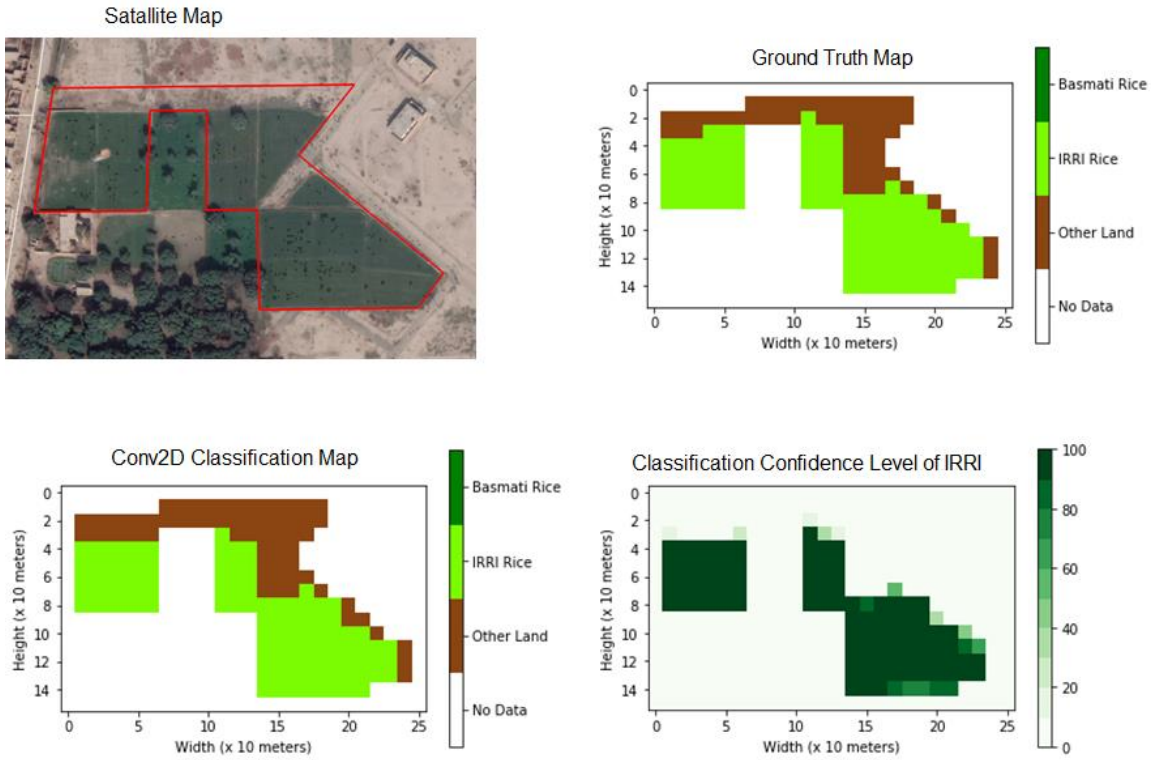

***Figure 46**: IRRI rice satellite, ground truth, classification, and confidence level map.*

The visual performance evaluation of the Conv2D is difficult, but overall the Basmati and IRRI Conv2D generated maps are similar to the ground truth map except for some pixels, also the confidence level map for both varieties are above 90% in most of the region.

6.3 Comparison With Previous Work

A comparison with the previous work in terms of overall accuracy are listed in the table below. The related work uses deep learning algorithms with different sensors dataset in different regions of the world to classify rice crops. Also, the authors adopted different strategies before input data to the classification model.

***Table 8**: Comparision of overall accuracies with the previous work.*

Author	Sensor	Method	Overall Accuracy
Yiqing Guo *et al*., in 2018	Sentinel-2	CNN	92.8%
Zhuang Zhou *et al*., in 2018	Sentinel-2	CNN	95.6%
Meng Zhang *et al.*, in 2018	Landsat-8	CNN+LST	97.8%
Weichun Zhang et al, in 2020	Sentinel-2	CNN	93.14%
Current Method	Sentinel-2	CNN	98.6%

Yiqing Guo and *et al*., in 2018, conduct a study in New South Wales, Australia, and constructed deep CNN to classify five rice varieties during the 2016–17 growing season and achieve an overall accuracy of 92.8%. Zhuang Zhou *et al*., in 2018 proposed a crop classification method based on CNN using Sentinel-2A multi-spectral images in Yuanyang county, Henan province, China, and achieve an overall accuracy of 95.6%. Meng Zhang *et al.*, in 2018 employed multitemporal Landsat 8 data and Patch-based deep-learning CNN with land-surface temperature (LST) to improve the accuracy of classification of paddy rice in the Dongting Lake of China and achieve a good overall accuracy of 97.8.

Weichun Zhang et al, in 2020 conduct experiments in Banan District and Zhongxian County Southwest hill area in China using Sentinel-2 multitemporal data for mapping rice paddy areas by developing four different CNN algorithms (Conv-1D, Conv-2D, two Conv-3D). For each classifier two schemes were designed: local classification and pre-trained classifier, and 2D-CNN performs best among all other with an overall accuracy of 93.14. In comparison to these, Conv2D by using high spectral and temporal resolutions and extracting the sub-pixel levels information with spectral unmixing gives the best overall accuracy and best suitable for rice crop varieties classification.

7. CONCLUSION AND FUTURE WORK

7.1 Conclusion

This research presents the classification of rice crop varieties Basmati and IRRI in Multan District, Punjab, Pakistan during growing season 2019 using Sentinel-2 multispectral and multitemporal images. In Pakistan where crop monitoring and management mostly relay on field surveys that consume a lot of resources and time, satellite imagery like Sentinel-2 has high potential to provide crop data accurately and timely for decision making and also for developing sustainable systems.

This section concludes the objectives of this research work which are presented in section 1.3.

1. *Review the existing deep learning approaches* utilized *for rice crop classification and select suitable architecture.*

From section 3.1.1 to 3.1.3 the existing approaches for the classification of crops with the focus on rice were reviewed and 2D CNN was selected for its performance among the other CNN architecture like 1D and 3D CNN. Additionally, CNN architecture performs way better than most popular classical machine learning approaches SVM and RF.

2. *Extract the sub-pixel level details from ground truth data using spectral unmixing methods.*

To fulfill that objective, Linear Mixing Model was adopted because it is most suitable for the area where endmembers are in a spatially discrete pattern like different crop fields. Subpixel level information was extracted and construct an abundance representation map of each rice variety.

3. *Extract the rice crop growing phenology details from vegetation indices.*

The growing habits of the rice varieties are studied from three vegetation and water indices NDVI, NDWI, and NDMI. NDVI was used to check the health and density of the vegetation whereas NDWI and NDMI were used to extract the water and moisture contents in the field and canopies.

4. *Implement the chosen architecture and optimized its performance.*

As explained earlier 2D CNN architecture was selected for rice varieties classification. Its performance was optimized by performing different experiments and tune the hyperparameters like batch size, filter size, kernel size, and dropout percentage to get the best model out of it.

5. *Evaluate the performance of classification using performance metrics and compare it with the existing work.*

The evaluation of Conv2D was based on accuracy, precision, recall, F1, and Kappa score statistics and gives the highest overall accuracy as compared to the existing work in the related field. From the results of the current and previous work, it is obvious that multispectral and multitemporal data significantly improved the classification accuracy. Additionally in that work, the extraction of sub-pixel information also helps in the classification of rice varieties and gives remarkable results.

7.2 Future Work

In that work, we put our best efforts in the improvement of the rice crop varieties classification but there is always an opportunity for the enhancement that needs to be explored. The following list is the suggestions for the future to improve crop classification and mapping.

1. More vegetation, water, and soil indices need to be explored to give valuable information to the farmers like the best time for seeding, irrigation, fertilization, and harvesting.
2. Temporal resolution can be increased by using image fusion techniques as the major drawback of the passive sensors are data unavailable due to cloud cover which can be overcome by using active sensors.
3. Land surface temperature data from the Sentinel-3 mission also helps in studying crop phenology and classification.
4. Sentinel-2 has a spatial resolution of 10m which can be improved to superresolution using spectral unmixing methods which give sub-pixel information and simulated annealing method.
5. ESA also increases the Sentinel family by launching Sentinel-2C and Sentinel-2D satellites in 2020/2021 with enhanced capabilities, which provide valuable information and also helpful to the agriculture and researcher community worldwide.

8. BIBLIOGRAPHY

[1] U. Nations. *Transforming our world: the 2030 Agenda for Sustainable Development*. Available: https://sdgs.un.org/2030agenda

[2] J. B. Adams and A. R. Gillespie, *Remote sensing of landscapes with spectral images: A physical modeling approach*. Cambridge University Press, 2006.

[3] J. Dong *et al.*, "Tracking the dynamics of paddy rice planting area in 1986–2010 through time series Landsat images and phenology-based algorithms," *Remote Sensing of Environment*, vol. 160, pp. 99-113, 2015.

[4] Y. Zhou *et al.*, "Mapping paddy rice planting area in rice-wetland coexistent areas through analysis of Landsat 8 OLI and MODIS images," *International journal of applied earth observation and geoinformation*, vol. 46, pp. 1-12, 2016.

[5] D. Peng, A. R. Huete, J. Huang, F. Wang, and H. Sun, "Detection and estimation of mixed paddy rice cropping patterns with MODIS data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 13, no. 1, pp. 13-23, 2011.

[6] M. Immitzer, F. Vuolo, and C. Atzberger, "First experience with Sentinel-2 data for crop and tree species classifications in central Europe," *Remote Sensing*, vol. 8, no. 3, p. 166, 2016.

[7] M. Drusch *et al.*, "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote sensing of Environment*, vol. 120, pp. 25-36, 2012.

[8] Q. Zhao *et al.*, "Investigating within-field variability of rice from high resolution satellite imagery in Qixing Farm County, Northeast China," *ISPRS International Journal of Geo-Information*, vol. 4, no. 1, pp. 236-261, 2015.

[9] Ç. Küçük, G. Taşkın, and E. Erten, "Paddy-rice phenology classification based on machine-learning methods using multitemporal co-polar X-band SAR images," *IEEE Journal of selected topics in applied earth observations and remote sensing*, vol. 9, no. 6, pp. 2509-2519, 2016.

[10] E. Madigan, Y. Guo, M. Pickering, A. Held, and X. Jia, "Quantitative monitoring of complete rice growing seasons using Sentinel 2 time series images," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 7699-7702: IEEE.

[11] F. D. Govt. of Pakistan, "PAKISTAN ECONOMIC SURVEY 2019-20."

[12] T. P. C. R. A. Limited. *Rice sector overview FY2020*. Available: https://www.pacra.com.pk/uploads/doc_report/Rice%20Sector%20Upload%20Version%20FY20_1573215111.pdf

[13] R. Houstoun, "Newton and the colours of the spectrum," *Science Progress (1916-1919)*, vol. 12, no. 46, pp. 250-264, 1917.

[14] P. Fara, "Newton shows the light: a commentary on Newton (1672)'A letter… containing his new theory about light and colours…'," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 373, no. 2039, p. 20140213, 2015.

[15] MIT. *The History of Spectroscopy, A Perspective*. Available: http://web.mit.edu/spectroscopy/history/spec-history.html

[16] N. S. t. B. Pettorelli, Henrike & Shapiro, Aurelie & Glover-Kapfer, Paul. (2018) SATELLITE REMOTE SENSING FOR CONSERVATION. *Conservation Technology Series Issue 4*.

[17] A. F. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for earth remote sensing," *science*, vol. 228, no. 4704, pp. 1147-1153, 1985.
[18] J. S. Pearlman, P. S. Barry, C. C. Segal, J. Shepanski, D. Beiso, and S. L. Carman, "Hyperion, a space-based imaging spectrometer," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1160-1173, 2003.
[19] M. Sivakumar, P. Roy, K. Harmsen, and S. Saha, *Satellite remote sensing and GIS applications in agricultural meteorology*.
[20] E. Shimizu, *SATELLITE REMOTE SENSING*.
[21] C. Toth and G. Jóźków, "Remote sensing platforms and sensors: A survey," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 115, pp. 22-36, 2016.
[22] topper. *Propagation of Electromagnetic Waves*. Available: https://www.toppr.com/guides/physics/communication-systems/propagation-of-electromagnetic-waves/
[23] S. Multimäki. (2019). *VISUALIZATION OF GEOGRAPHICAL INFORMATION*. Available: https://mycourses.aalto.fi/pluginfile.php/1150624/mod_resource/content/1/Mon%202.12..pdf
[24] J. McLaughlin and K. Webster, *Effects of a changing climate on peatlands in permafrost zones: a literature review and application to Ontario's Far North* (no. CCRR-34). Ontario Forest Research Institute, 2013.
[25] G. Geography. (2020). *Spectral Signature Cheatsheet – Spectral Bands in Remote Sensing*. Available: https://gisgeography.com/spectral-signature/
[26] P. KNOWLEDGE. *Specular Reflection vs. Diffuse Reflection*. Available: http://physics-knowledge.blogspot.com/2010/05/specular-raflection-vs-diffuse.html
[27] seos, "Introduction to Categorisation of Objects from their Data."
[28] J. Rouse Jr, R. Haas, J. Schell, and D. Deering, "Paper A 20," in *Third Earth Resources Technology Satellite-1 Symposium: The Proceedings of a Symposium Held by Goddard Space Flight Center at Washington, DC on December 10-14, 1973: Prepared at Goddard Space Flight Center*, 1974, vol. 351, p. 309: Scientific and Technical Information Office, National Aeronautics and Space ….
[29] A. Huete, "Huete, AR A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment," *Remote sensing of environment*, vol. 25, pp. 295-309, 1988.
[30] B.-C. Gao, "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space," *Remote sensing of environment*, vol. 58, no. 3, pp. 257-266, 1996.
[31] M. Hardisky, R. Smart, and V. Klemas, "Seasonal spectral characteristics and aboveground biomass of the tidal marsh plant," *Spartina-Alterniflora Photogrammetric Engineering and Remote Sensing*, vol. 49, pp. 85-92, 1983.
[32] F. K. Lutgens, E. J. Tarbuck, and D. Tusa, *The atmosphere*. Prentice-Hall Englewood Cliffs, NJ, 1995.
[33] E. Portal. *spaceborne EO missions and sensors*. Available: https://directory.eoportal.org/web/eoportal/satellite-missions
[34] ESA. *Copernicus European's eyes on earth*. Available: https://www.copernicus.eu/en
[35] D. Nakar, "Sentinel-2: Multispectral Instrument (MSI) design and system performance," 2019.
[36] ESA. (2015). *Sentinel-2 User Handbook*. Available: https://earth.esa.int/documents/247904/685211/Sentinel-2_User_Handbook

[37] F. Gascon *et al.*, "Copernicus Sentinel-2A calibration and products validation status," *Remote Sensing*, vol. 9, no. 6, p. 584, 2017.
[38] S. Vieira, W. H. Pinaya, and A. Mechelli, "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications," *Neuroscience & Biobehavioral Reviews*, vol. 74, pp. 58-75, 2017.
[39] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
[40] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436-444, 2015.
[41] A. Mechelli and S. Vieira, *Machine Learning: Methods and Applications to Brain Disorders*. Academic Press, 2019.
[42] K. Hara, D. Saito, and H. Shouno, "Analysis of function of rectified linear unit used in deep learning," in *2015 International Joint Conference on Neural Networks (IJCNN)*, 2015, pp. 1-8: IEEE.
[43] S. J. Nowlan and G. E. Hinton, "Simplifying neural networks by soft weight-sharing," *Neural computation*, vol. 4, no. 4, pp. 473-493, 1992.
[44] L. Prechelt, "Automatic early stopping using cross validation: quantifying the criteria," *Neural Networks*, vol. 11, no. 4, pp. 761-767, 1998.
[45] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929-1958, 2014.
[46] D. R. Wilson and T. R. Martinez, "The need for small learning rates on large problems," in *IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No. 01CH37222)*, 2001, vol. 1, pp. 115-119: IEEE.
[47] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Neural networks for perception*: Elsevier, 1992, pp. 65-93.
[48] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533-536, 1986.
[49] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[50] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural networks: Tricks of the trade*: Springer, 2012, pp. 437-478.
[51] M. U. Liaqat, M. J. M. Cheema, W. Huang, T. Mahmood, M. Zaman, and M. M. Khan, "Evaluation of MODIS and Landsat multiband vegetation indices used for wheat yield estimation in irrigated Indus Basin," *Computers and Electronics in Agriculture*, vol. 138, pp. 39-47, 2017.
[52] H. Qiong *et al.*, "How do temporal and spectral features matter in crop classification in Heilongjiang Province, China?," *Journal of Integrative Agriculture*, vol. 16, no. 2, pp. 324-336, 2017.
[53] A. ULLAH and S. ZAFAR, "MAPPING CROPS FROM THEIR TEMPORAL BEHAVIOR BY INTEGRATING SENTINEL-2 SATELLITE AND GROUND DATA," *FUUAST Journal of Biology*, vol. 8, no. 2, pp. 217-224, 2018.
[54] A. Karamat *et al.*, "Estimation of Net Rice Production for the Fiscal year 2019 using Multisource Datasets," *Int. J. Agric. Sustainable Dev*, vol. 1, no. 02, pp. 47-65, 2019.
[55] M. Singha, B. Wu, and M. Zhang, "An object-based paddy rice classification using multi-spectral data and crop phenology in Assam, Northeast India," *Remote Sensing*, vol. 8, no. 6, p. 479, 2016.

[56] Y. Guo, X. Jia, and D. Paull, "Sequential classifier training for rice mapping with multitemporal remote sensing imagery," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* vol. 4, p. 161, 2017.

[57] R. Saini and S. Ghosh, "CROP CLASSIFICATION ON SINGLE DATE SENTINEL-2 IMAGERY USING RANDOM FOREST AND SUPPOR VECTOR MACHINE," *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences,* 2018.

[58] A. Khan *et al.*, "Evaluating Landsat and Rapideye data for winter wheat mapping and area estimation in Punjab, Pakistan," *Remote Sensing,* vol. 10, no. 4, p. 489, 2018.

[59] S. Feng, J. Zhao, T. Liu, H. Zhang, Z. Zhang, and X. Guo, "Crop type identification and mapping using machine learning algorithms and sentinel-2 time series data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 12, no. 9, pp. 3295-3306, 2019.

[60] L. Zhao, Y. Shi, B. Liu, C. Hovis, Y. Duan, and Z. Shi, "Finer Classification of Crops by Fusing UAV Images and Sentinel-2A Data," *Remote Sensing,* vol. 11, no. 24, p. 3012, 2019.

[61] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geoscience and Remote Sensing Letters,* vol. 14, no. 5, pp. 778-782, 2017.

[62] Z. Zhou, S. Li, and Y. Shao, "Crops Classification from Sentinel-2A Multi-spectral Remote Sensing Images Based on Convolutional Neural Networks," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 5300-5303: IEEE.

[63] Y. Guo, X. Jia, and D. Paull, "Mapping of rice varieties with sentinel-2 data via deep cnn learning in spectral and time domains," in *2018 Digital Image Computing: Techniques and Applications (DICTA)*, 2018, pp. 1-7: IEEE.

[64] S. Ji, C. Zhang, A. Xu, Y. Shi, and Y. Duan, "3D convolutional neural networks for crop classification with multi-temporal remote sensing images," *Remote Sensing,* vol. 10, no. 1, p. 75, 2018.

[65] M. Zhang, H. Lin, G. Wang, H. Sun, and J. Fu, "Mapping paddy rice using a convolutional neural network (CNN) with Landsat 8 datasets in the Dongting Lake Area, China," *Remote Sensing,* vol. 10, no. 11, p. 1840, 2018.

[66] Z. Sun, L. Di, H. Fang, and A. Burgess, "Deep Learning Classification for Crop Types in North Dakota," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 2020.

[67] W. Zhang, H. Liu, W. Wu, L. Zhan, and J. Wei, "Mapping Rice Paddy Based on Machine Learning with Sentinel-2 Multi-Temporal Data: Model Comparison and Transferability," *Remote Sensing,* vol. 12, no. 10, p. 1620, 2020.

[68] A. Villa, J. Chanussot, J. A. Benediktsson, and C. Jutten, "Spectral unmixing for the classification of hyperspectral images at a finer spatial resolution," *IEEE Journal of Selected Topics in Signal Processing,* vol. 5, no. 3, pp. 521-533, 2010.

[69] I. Dópido and A. Plaza, "Unmixing prior to supervised classification of urban hyperspectral images," in *2011 Joint Urban Remote Sensing Event*, 2011, pp. 97-100: IEEE.

[70] E. A. Mylona, O. A. Sykioti, K. D. Koutroumbas, and A. A. Rontogiannis, "Spectral unmixing-based clustering of high-spatial resolution hyperspectral imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 10, no. 8, pp. 3711-3721, 2017.

[71] E. Mylona, V. Daskalopoulou, O. Sykioti, K. Koutroumbas, and A. Rontogiannis, "Classification of sentinel-2 images utilizing abundance representation," in *Multidisciplinary Digital Publishing Institute Proceedings*, 2018, vol. 2, no. 7, p. 328.
[72] S. Xu, J. Li, M. Khodadadzadeh, A. Marinoni, P. Gamba, and B. Li, "Abundance-indicated subspace for hyperspectral classification with limited training samples," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 4, pp. 1265-1278, 2019.
[73] B. Fang, Y. Bai, and Y. Li, "Combining spectral unmixing and 3d/2d dense networks with early-exiting strategy for hyperspectral image classification," *Remote Sensing*, vol. 12, no. 5, p. 779, 2020.
[74] ESA. *Copernicus Open Access Hub*. Available: https://scihub.copernicus.eu/
[75] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE signal processing magazine*, vol. 19, no. 1, pp. 44-57, 2002.
[76] J. B. Adams, M. O. Smith, and P. E. Johnson, "Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 site," *Journal of Geophysical Research: Solid Earth*, vol. 91, no. B8, pp. 8098-8112, 1986.
[77] G. Storvik, R. Fjortoft, and A. H. S. Solberg, "A Bayesian approach to classification of multiresolution remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 539-547, 2005.
[78] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2014-2039, 2011.
[79] R. P. P. Machado and C. Small, "Identifying multi-decadal changes of the Sao Paulo urban agglomeration with mixed remote sensing techniques: Spectral mixture analysis and night lights," *EARSeL eProceedings*, vol. 12, no. 2, p. 101, 2013.
[80] L. C. Parra, C. Spence, P. Sajda, A. Ziehe, and K.-R. Müller, "Unmixing hyperspectral data," in *Advances in neural information processing systems*, 2000, pp. 942-948.
[81] D. Lu, M. Batistella, and E. Moran, "Multitemporal spectral mixture analysis for Amazonian land-cover change detection," *Canadian Journal of Remote Sensing*, vol. 30, no. 1, pp. 87-100, 2004.
[82] D. C. Heinz, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE transactions on geoscience and remote sensing*, vol. 39, no. 3, pp. 529-545, 2001.
[83] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[84] P. M. Radiuk, "Impact of training set batch size on the performance of convolutional neural networks for diverse datasets," *Information Technology and Management Science*, vol. 20, no. 1, pp. 20-24, 2017.
[85] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, 2018.
[86] Y. Sasaki, "The truth of the f-measure. 2007," ed, 2007.
[87] J. Sim and C. C. Wright, "The kappa statistic in reliability studies: use, interpretation, and sample size requirements," *Physical therapy*, vol. 85, no. 3, pp. 257-268, 2005.

9. APPENDICES

9.1 Appendix A: ESA SNAP Software

That section provides details about the SNAP software interface and processing steps that were followed in the research work.

Overview of interface

The SNAP software main interface is divided into several parts and each detail is given below.

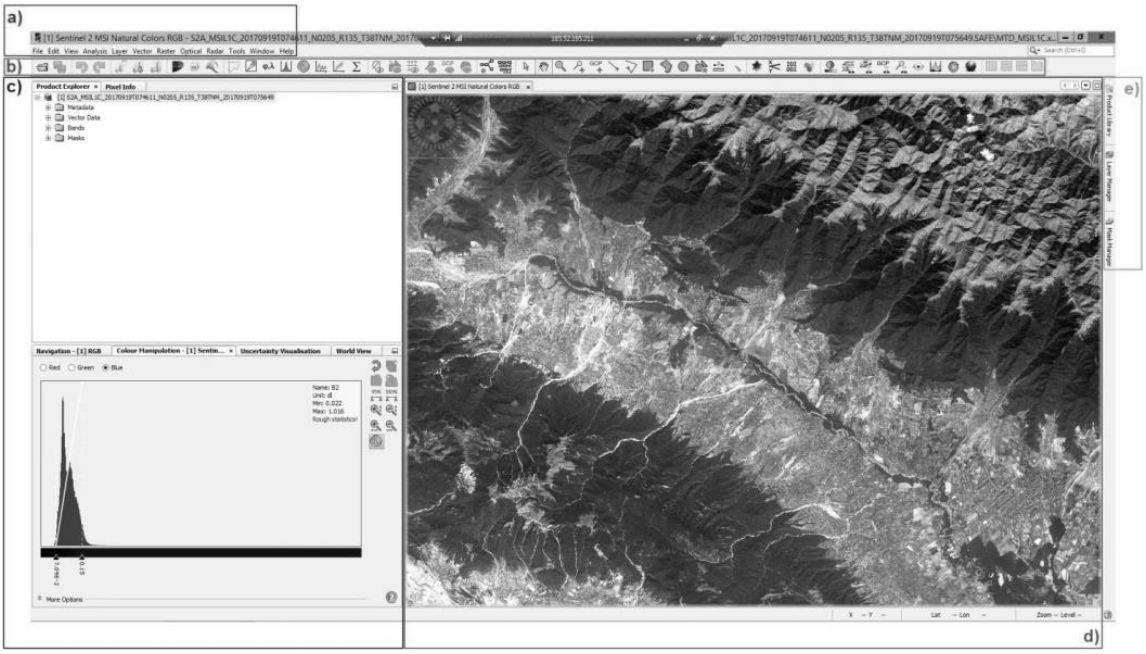

***Figure 47**: ESA SNAP main window.*

a. Main toolbar – used for reading, writing, process data, and application configuration.
b. Shortcut toolbar – direct links to software functions.
c. Navigation and data display panel – used for easy navigation, change visualization, and editing functions.
d. Product library and layers manager panel – manage image dataset and different layers presentation on the map view.
e. Map view – visualization of images.

Read, View and Save Data

Reading or importing in SNAP

Step 1. Open file adding window from the main toolbar File → Open Product or from shortcut icon.

Step 2. In the Open Product window select Sentinel-2 MSI product or tile (*.xml) from the drop-down list of Files of type.

Step 3. Find the Sentinel-2 data directory, open the main catalog, and select MTD_MSIL2A.xml file and click Open, the image will be added to Product Explorer.

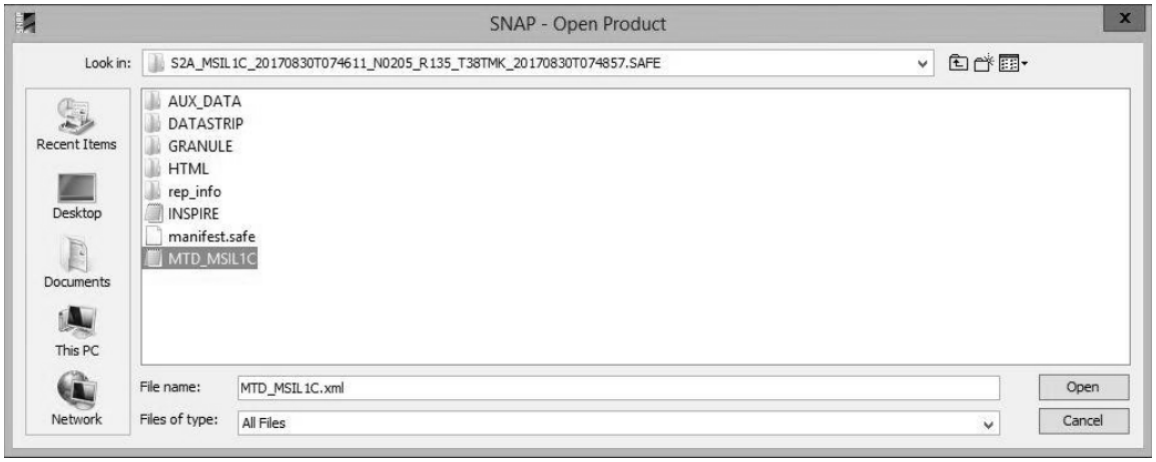

***Figure 48**: Open product in ESA SNAP.*

Display image in SNAP

Step 1. Open the map window tool from the main toolbar by Window → Open RGB Image Window or by right click on the added file in product explorer and select Open RGB Image Window.

Step 2. In the open window set Profile and RGB channels and click OK.

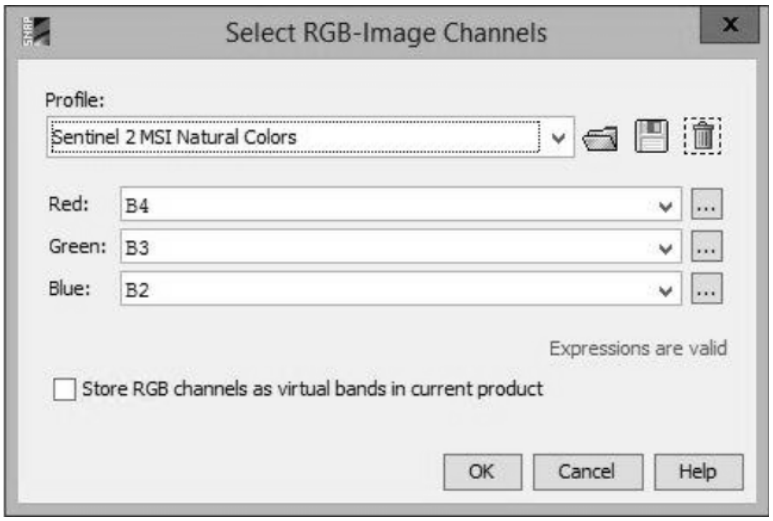

***Figure 49**: Open RGB image window.*

Saving Data in SNAP

Data saving in ESA SNAP can be done in different ways which are described below.

a. If changes save in the same file the open from the main toolbar by File → Save Product.
b. If save a copy of changes then File → Save Product As.
c. If data format change is required then File → Export.

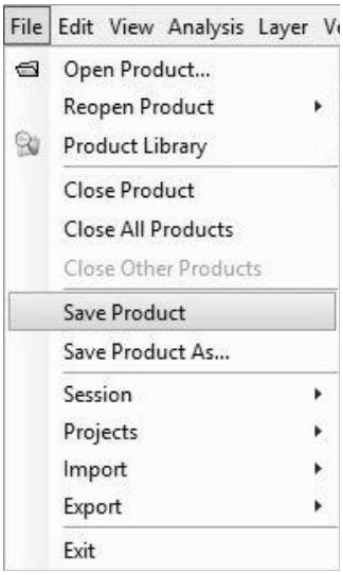

***Figure 50**: Data save options in ESA SNAP.*

Resampling

Step 1. Running the resampling window in Raster → Geometric Operations → Resampling.

Step 2. In the I/O Parameter tab select the source product from the dropdown list and the target product name is set by default to the source file name with a resampled suffix.

Step 3. In the Resampling Parameter tab define the size of the resampled product to By reference band from source product option and select a band with a spatial resolution of 10m. leave the other settings as default and run the resampling process by pressing the Run button. All the bands with different spatial resolution will be resampled to 10m resolution.

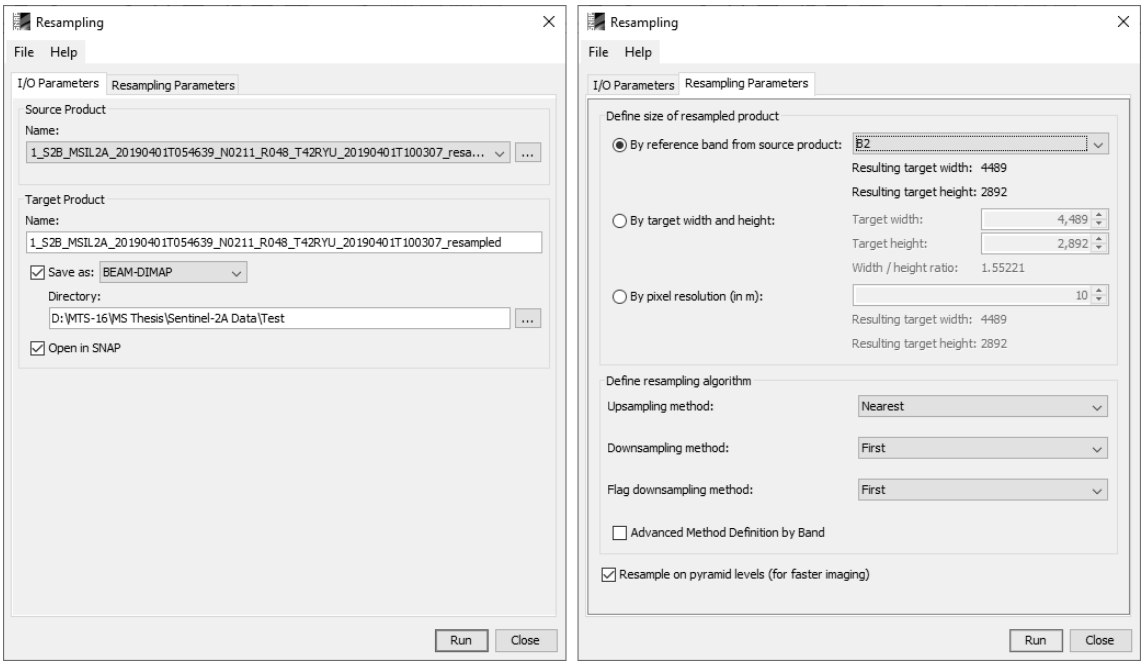

***Figure 51**: I/O and resampling parameters settings window.*

Reprojection

Step 1. run from the main toolbar by Raster → Geometric Operations → Reprojection.

Step 2. In the I/O Parameter tab select the source product from the dropdown list and the target product name is set by default to the source file name with a reprojected suffix.

Step 3. In reprojection parameters, tab by default custom CRS with Geographic Lat/Long WGS84 projection was set, and if the user wants to change the CRS then select the required projection and leave all other settings unchanged and run reprojection process from Run button.

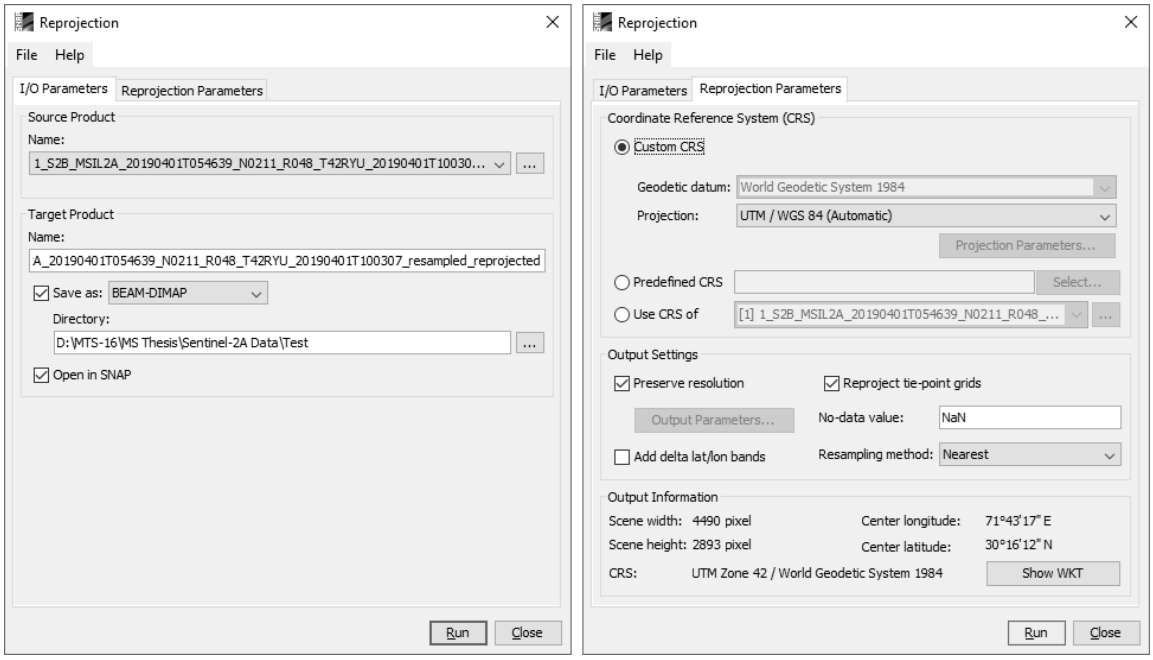

***Figure 52**: I/O and reprojection parameter setting window.*

Clipping and Masking AOI

Step 1. Run the subset window from the main toolbar by Raster → Subset.

Step 2. In the Pixel Coordinates method enter the exact pixel location of the staring and ending subset, and in Geo-Coordinates method Latitude and Longitude bound values are entered in decimal notation.

Step 3. In the Band Subset tab select the bands which are necessary for further processing and analysis was selected and click the Ok button. The target product name is set by default to the source file name with a subset prefix.

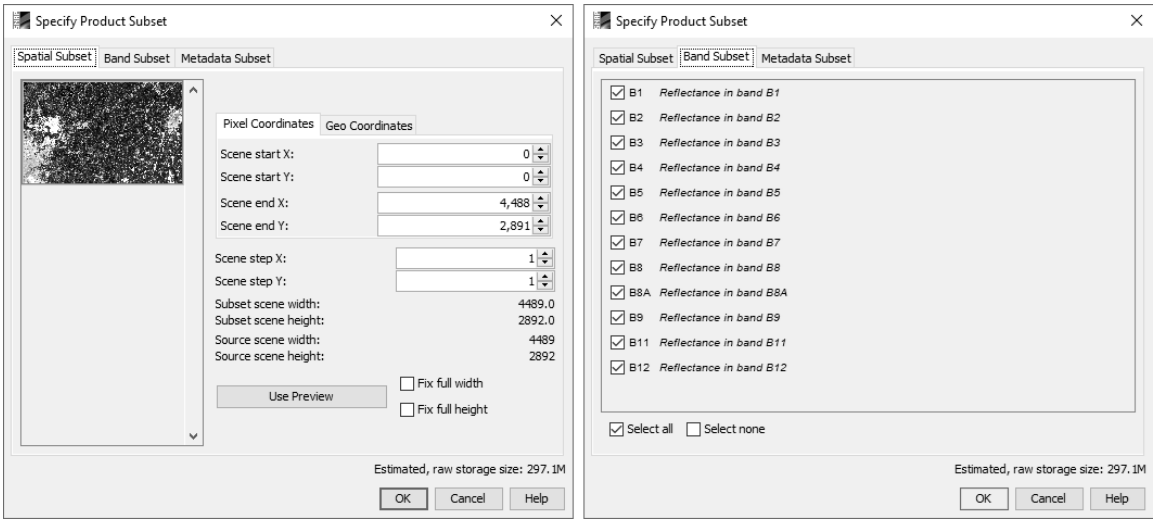

***Figure 53**: Spatial subset and Band subset window.*

Step 4. The clipped product will be added to the product explorer automatically. Select the product and open vector shapefile window of the area of interest form the main toolbar by Vector → Import → ESRI Shapefile. (ESRI shapefile are created in QGIS form KMZ polygon file of the area of interest.)

Step 5. Run the Land Sea Mask tool from the main toolbar by Raster → Mask → Land/Sea Mask.

Step 6. In the I/O Parameter tab select the source product from the dropdown list and the target product name is set by default to the source file name with a msk suffix.

Step 7. In the processing parameters tab select the bands, select the Use SRTM 3 sec, and select the Use Vector as Mask and select shapefile from the drop-down list and click Run button.

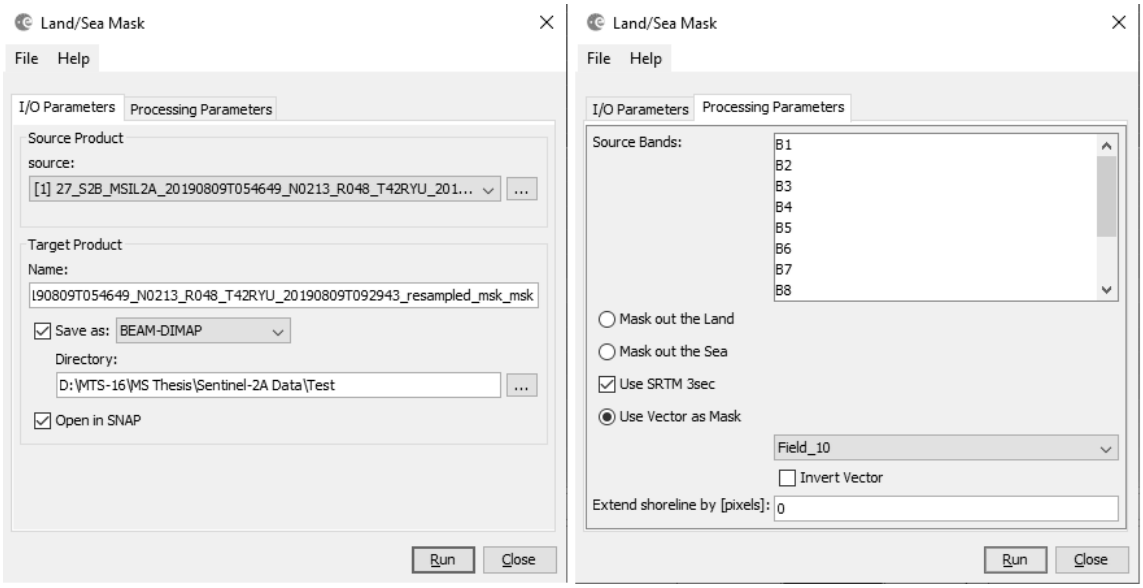

***Figure 54**: Land/Sea Mask I/O and processing parameters window.*

Spectral indices

Step 1. Run the vegetation indices processor from the main toolbar by Optical → Thematic Land Processing → Vegetation Radiometric Indices → NDVI Processor. To run water indices processor Optical → Thematic Land Processing → Water Radiometric Indices → NDWI Processor.

Step 2. In the I/O Parameter tab select the source product from the dropdown list and the target product name is set by default to the source file name with an index name (ndvi, ndwi, etc.) suffix.

Step 3. In the Processing Parameter tab select the bands used in a particular index and its factor and click run.

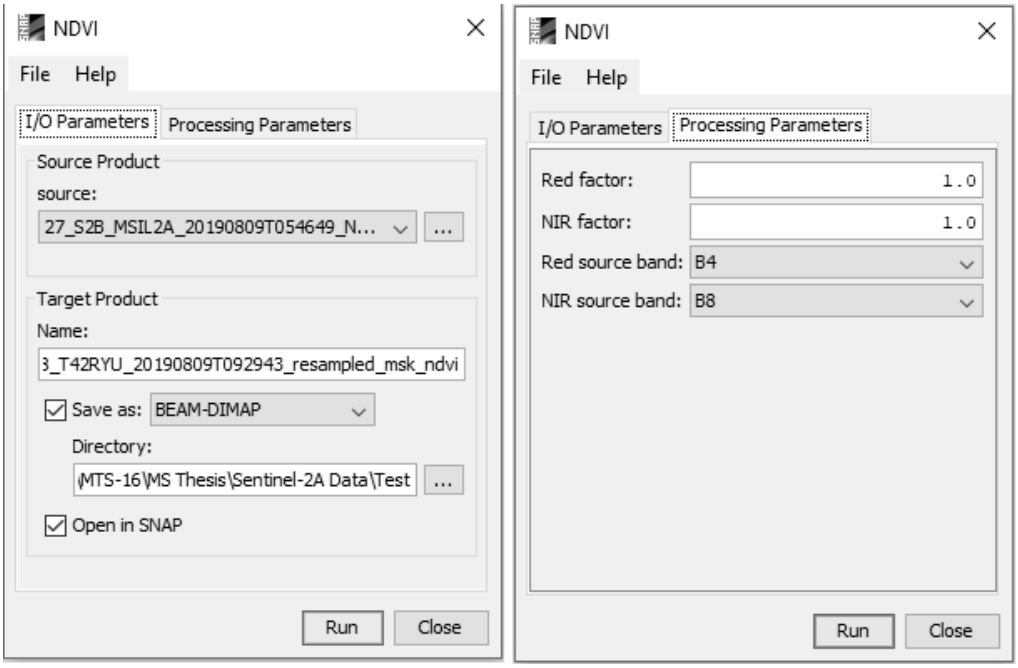

***Figure 55**: Vegetation index NDVI processing window.*

Step 4. Spectral indices can also be calculated using the Band Math tool from the main toolbar by Raster Band Maths and edit expression for a particular index.

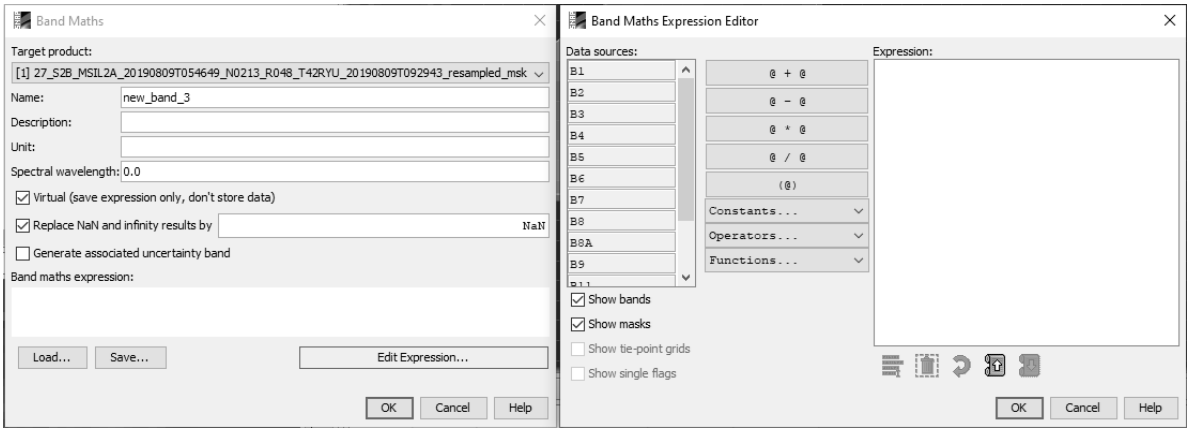

***Figure 56**: Band Math and Expression Editor window.*

Spectral Unmixing

Step 1. Open the image in the RGB window and select the pure pixel using the Pin Manager tool form the main tollbar by View → Tool Windows → Pin Manager.

Step 2. Import the pure pixel location in the Pin Manager window in XML or text format as shown in **Figure 35**.

Step 3. Open Spectrum View from the main toolbar by Optical Spectrum View and export the spectra of endmembers to a text file.

Step 4. Open Spectral Unmixing window from the main toolbar by Optical Spectral Unmixing and select the endmember spectra which we save in step 3, also select bands and their minimum bandwidth, unmixing model, and click Run button.

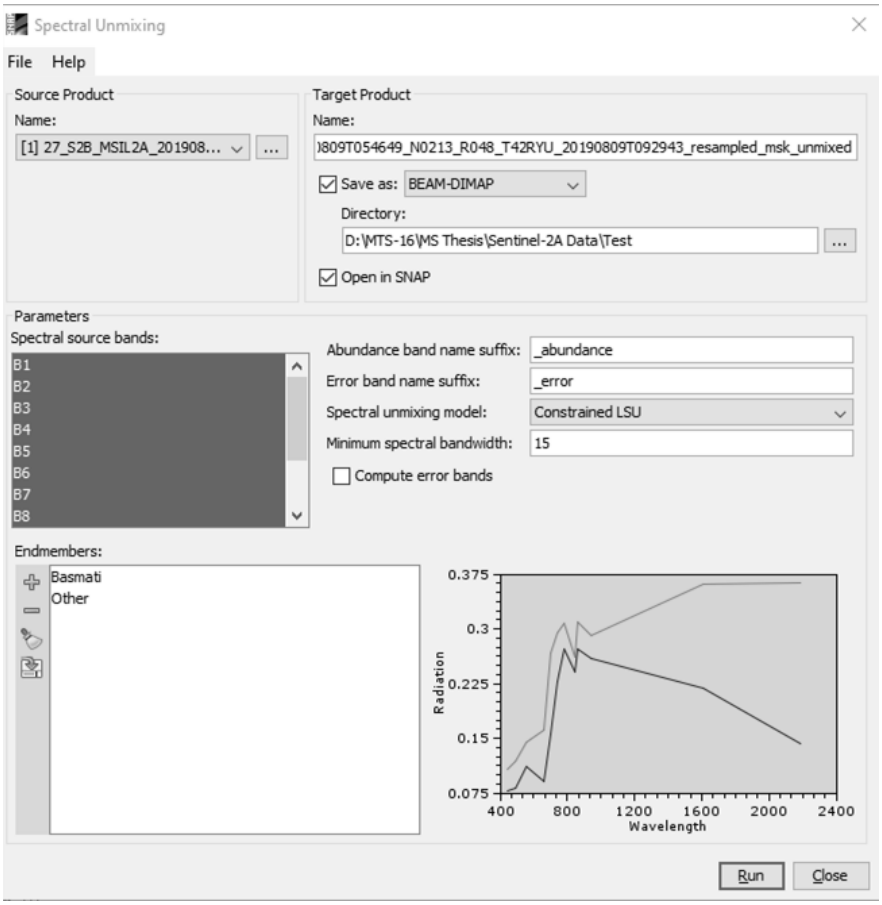

***Figure 57**: Spectral Unmixing parameter setting window.*

Graph Builder

Almost all the data processing steps which are discussed earlier will be processed automatically in a defined series and generate the final output data using the Graph Builder tool in ESA SANP software. The following steps were practices to generate the graph, process data, and finally save the product in the desired file format.

Step 1. Run graph builder tool from the main toolbar by Tools → GraphBuilder.

Step 2. In the GraphBuilder window right-click the mouse and add the required processing steps in the defined sequence and connect them all.

Step 3. Select each process parameter tab and set the parameters' values as set in the individual processing steps and click the Run button. After all the processing is done the final product will be open in the product explorer window.

***Figure 58**: SNAP GraphBuilder window.*