

**Sugarcane Crop Classification from Satellite Imagery  
using Deep Learning Approach**



**Author**

**Sidra Muquddas**

**00000319775**

**Supervised by**

**Prof. Dr. Hamid Jabbar**

**MASTERS IN MECHATRONICS ENGINEERING,**

**DEPARTMENT OF MECHATRONICS ENGINEERING,  
COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING,  
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,  
ISLAMABAD, PAKISTAN.**

**JULY 2023**

## THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis by Ms. Sidra Muquddas Registration No. 00000319775, of Electrical and Mechanical Engineering College has been vetted by undersigned, found complete in all respects as per NUST Statues/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature:  \_\_\_\_\_

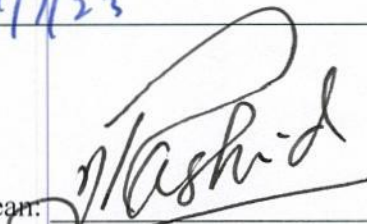
Name of Supervisor: Dr. Hamid Jabbar

Dated: 19/7/23 \_\_\_\_\_

Signature of HOD:  \_\_\_\_\_

Dr. Amir Hamza

Date: 19/7/23 \_\_\_\_\_

Signature of Dean:  \_\_\_\_\_

Brig Dr. Nasir Rashid

Date: 19 JUL 2023 \_\_\_\_\_

*With unwavering support and encouragement from my exceptional parents, beloved husband, and adored siblings, I have achieved this remarkable accomplishment.*

## **Acknowledgments**

I am grateful to Almighty Allah, the most Merciful for giving me the ability to accomplish my thesis.

I am grateful to my cherished guardians, my beloved husband, and my adored siblings for their help to me in every period of my life.

I might likewise want to communicate unique gratitude to my supervisor Dr.Hamid Jabbar and Dr. Waqar Shahid for their assistance throughout the duration of my thesis. I can confidently affirm that without their assistance, the completion of this research would not have been possible.

I would also like to extend my heartfelt gratitude to my classmate Ramsha Shahid for her endless support and motivation. Her willingness to share knowledge fostered a supportive learning environment.

I am also thankful to Ayaz Fayyaz and all staff of the Mechatronics Engineering department for their support and cooperation.

## Abstract

Sugarcane holds great importance as a crop in Pakistan, being one of the top global producers. Precise mapping of sugarcane fields is vital for effectively monitoring their size, production, and evaluating their influence on society, the economy, and the environment. This study focuses on presenting a deep learning-based framework that employs pixel-based classification to identify sugarcane crops in Pakistan, along with other commonly grown crops, using Sentinel 2A multi-spectral imagery. The framework encompasses various stages, including the selection of Sentinel products, preprocessing, extraction of spectral indices, a compilation of spectral features, labeling through spectral unmixing, and crop classification. The selection of Sentinel products for each crop field is based on the NDVI values. Multiple spectral and biophysical indices are derived from these Sentinel products. Each pixel is compiled into an image, representing the temporal evolution of that pixel across spectral bands and indices. These synthetic images are resampled using bilinear sampling and fed into various deep-learning models to solve the classification problem. Additionally, linear spectral unmixing is employed to assign labels to each pixel, ensuring accurate identification of the predominant crop in that pixel. The dataset used in this study comprises samples from different districts in Pakistan. Two combinations of datasets are created to assess the robustness of the developed methodology: one for training and testing within the same district and another for training and testing across separate districts. For the first dataset combination, most of the tested classification models achieve a high accuracy of approximately 99%. In the second dataset combination, LSTM outperforms other models, achieving an accuracy of 90%. The classified pixels are then integrated into the classification map, accurately representing their geolocation within the corresponding field. The proposed framework demonstrates promising results compared to the convNext model and exhibits the potential to effectively classify ready for harvest sugarcane among other crops using a limited number of products. Furthermore, it proves capable of classifying sugarcane across different districts.

**Keywords:** *Convolution neural network (CNN), Sentinel-2, Spectral Un-mixing, Long short-term memory (LSTM), Normalized difference vegetation index (NDVI), Deep Learning, Sugarcane*

# Table of Content

THESIS ACCEPTANCE CERTIFICATE.....	II
Acknowledgments .....	IV
Abstract.....	VI
Table of Content .....	VII
List of Figures.....	IX
List of Tables .....	XI
List Abbreviations .....	XII
Chapter 1: Introduction.....	1
1.1 Overview .....	1
1.2 Contribution of Sugarcane to Pakistan’s Economy .....	1
1.3 Challenges in Sugarcane Production and Management .....	3
1.4 Satellites used in Crop Classification .....	4
1.5 Research Objectives .....	4
Chapter 2: Literature Review.....	7
2.1 Crop mapping using conventional machine learning techniques .....	7
2.1.1 Incorporating single-date data .....	7
2.1.2 Incorporating time-series data .....	8
2.2 Crop Mapping using time weighted dynamic time warping method .....	9
2.3 Crop mapping using deep learning techniques.....	10
2.3.1 Using Recurrent Neural Network and LSTM.....	11
2.3.2 Using Convolutional Neural Network.....	12
2.3.3 Convolution Neural Network with Attention Layers .....	12
2.4 Research Gap Analysis .....	13
Chapter 3: Study Area and Materials .....	15
3.1 Study Area .....	15
3.2 Spectral Indices.....	16
3.3 Spectral Unmixing.....	17
3.4 Classification Models .....	18
3.4.1 VGG-16 .....	19
3.4.2 ResNet-50 .....	19

3.4.3 Inception v3 .....	20
3.4.4 Long Short-term Memory .....	20
3.4.5 ConvNext .....	21
Chapter 4: Proposed Framework .....	23
4.1 Product Selection .....	23
4.2 Preprocessing .....	26
4.3 Feature Extraction .....	28
4.4 Compilation of 2D spectral feature maps .....	29
4.5 Labeling .....	29
4.6 Dataset combination .....	30
4.7 Training Classification Models .....	30
Chapter 5: Results & Discussion .....	34
5.1 Proposed Framework for Sugarcane classification .....	34
5.1.1 Confusion matrix .....	34
5.1.2 Classification results for dataset combination 1 .....	36
5.1.3 Classification results for dataset combination 2 .....	39
Chapter 6: Conclusion .....	43
BIBLIOGRAPHY .....	45

## List of Figures

<b>Figure 1.1:</b> Flowchart of the Research.....	6
<b>Figure 3.1:</b> Study area.....	15
<b>Figure 3.2:</b> Spectral Unmixing[26].....	18
<b>Figure 3.3:</b> VGG-16 Architecture [29] .....	19
<b>Figure 3.4:</b> ResNet-50 Architecture [31] .....	20
<b>Figure 3.5:</b> Inception v3 Architecture[33] .....	20
<b>Figure 3.6:</b> Long short-term memory Recurrent Unit .....	21
<b>Figure 3.7:</b> ConvNext architecture[40].....	22
<b>Figure 4.1:</b> Block diagram of proposed framework for crop classification.....	23
<b>Figure 4.2:</b> Detailed Abstract of proposed framework .....	24
<b>Figure 4.3:</b> Crop calendar of investigated crops .....	25
<b>Figure 4.4:</b> Selection of time window for downloading sentinel products for a) sugarcane, b) wheat, c) rice, d) corn .....	26
<b>Figure 4.5:</b> Polygons used for clipping the fields of a) sugarcane, b) wheat, c) rice, d) corn from sentinel product in SNAP.....	28
<b>Figure 4.6:</b> Spectral features extracted from 12 spectral bands of sentinel image 28	
<b>Figure 4.7:</b> Compilation of time-series spectral features from field image having 12 spectral bands to 32x32x3 image .....	29
<b>Figure 4.8:</b> a) Endmember selection, b) Endmember spectrum extraction, c) abundance map of green vegetation .....	30
<b>Figure 4.9:</b> Classification models trained on sample size 10x18 .....	31
<b>Figure 4.10:</b> Classification models trained on sample size 32x32x3 .....	32
<b>Figure 4.11:</b> Classification models trained on sample size 75x75x3 .....	33
<b>Figure 5.1:</b> Confusion matrix[43] .....	34
<b>Figure 5.2:</b> Confusion matrix for the models with input sample size 10x18	36
<b>Figure 5.3:</b> Classification model MLP; a) cropped sugarcane field b) abundance map c) classification map green pixel shows sugarcane grey shows non-sugarcane, d) non-sugarcane field, e) abundance map of green vegetation, f) classification map green shows non-sugarcane and grey shows sugarcane.....	37



**Figure 5.4:** Classification model Convnet; a) cropped sugarcane field b) abundance map c) classification map green pixel shows sugarcane grey shows non- sugarcane, d) non-sugarcane field, e) abundance map of green vegetation, f) classification map green shows non-sugarcane and grey shows sugarcane..... 38

**Figure 5.5:** Classification model LSTM; a) cropped sugarcane field b) abundance map c) classification map green pixel shows sugarcane grey shows non-sugarcane, d) non-sugarcane field, e) abundance map of green vegetation, f) classification map green shows non-sugarcane and grey shows sugarcane..... 38

**Figure 5.6:** Confusion matrix for classification models VGG16, Resnet50, Inception v3..... 39

**Figure 5.9:** Classification model VGG16 for dataset combination 2; a) cropped sugarcane field b) abundance map c) classification map green pixel shows sugarcane grey shows non-sugarcane, d) non-sugarcane field, e) abundance map of green vegetation, f) classification map green shows non-sugarcane and grey shows sugarcane ..... 42

## List of Tables

Table 1.1: Sugarcane Area and Production .....	2
Table 1.2: Sugarcane Area and Production with respect to Province.....	3
Table 1.3: Satellite Comparison .....	4
Table 3.1: Spectral indices ranges .....	17
Table 4.1: Crop type and estimated crop life.....	27
Table 5.1: Evaluation metrics along with input sample size for dataset combination 1 .....	36
Table 5.2: Training and test accuracy for dataset combination 2.....	40
Table 5.3: Evaluation metrics along with input sample size for dataset combination 2 .....	40

## **List Abbreviations**

<b>CNN</b>	Convolution Neural Network
<b>Convnet</b>	Convolutional Network
<b>DL</b>	Deep Learning
<b>DVI</b>	Difference Vegetation Index
<b>FVC</b>	Fractional Vegetation Cover
<b>GIS</b>	Graphic Information System
<b>GPS</b>	Global Positioning System
<b>LAI</b>	Leaf Area Index
<b>LSTM</b>	Long Short-Term Memory
<b>ML</b>	Machine Learning
<b>MLP</b>	Multi-Layer Perceptron
<b>NCRA</b>	National center for Robotics and Automation
<b>NDVI</b>	Normalize Differential Vegetation Index
<b>NDWI</b>	Normalize Differential Water Index
<b>QGIS</b>	Quantum Geographic Information System

<b>RF</b>	Random Forest
<b>SAVI</b>	Soil Adjusted Vegetation Index
<b>SNAP</b>	Sentinel Application Platform
<b>SVM</b>	Support Vector Machines

# Chapter 1: Introduction

## 1.1 Overview

Agriculture is the backbone of the world's economy and in order to manage the resources for high-quality production of crops, an updated and well-maintained management system is essential. In order to get maximum productivity and high-quality products it is required to adapt modern methods that are observed worldwide for monitoring and management of agriculture resources. Traditional field surveys for crop monitoring requires a lot of man power resulting in time consumption, extravagant budgets and still the results are not accurate.

Precision agriculture plays a crucial role in crop identification, growth monitoring, yield estimation, and irrigation and soil management. Traditional methods like ground surveys and aerial imagery for crop monitoring are labor-intensive, time-consuming, and costly. However, the introduction of remote sensing technology in agriculture, starting with Landsat-1 in 1972, has revolutionized the monitoring and identification of crop types and growth stages, making it more cost-effective and efficient. Crop-type mapping offers valuable insights into crop yields and their economic significance in the agricultural sector. The European Space Agency's launch of Sentinel-2A in June 2015 followed by Sentinel-2B in March 2017 aimed to enhance spectral, spatial, and temporal resolution compared to previous satellites. These satellites provide multi-spectral imagery covering extensive land areas, islands, and coastal waters. When combined with machine learning and deep learning techniques, they enable effective monitoring of land cover without extensive on-site visits.

This study proposes a framework for the classification and mapping of sugarcane crop among various other popular crops. The proposed framework consists of selection of sentinel products based on time-series NDVI plots, along with generation of 2D time series spectral feature maps representing individual pixels, which are classified using multiple CNN architectures and LSTM and then fitting each pixel in classification maps.

## 1.2 Contribution of Sugarcane to Pakistan's Economy

The agricultural sector plays a critical role in Pakistan's economy, accounting for 24% of the country's GDP [1]. Additionally, nearly 30% of the labor workforce in Pakistan is associated with agricultural activities. However, this contribution has been declining in recent years due to various factors such as low productivity, climate change, inadequate infrastructure, and the prevalence of pests and diseases. These challenges collectively lead to reduced yields and hinder the sector's growth. One significant reason for low productivity is the failure to adopt modern monitoring and production farming techniques practiced in developed countries, as there is a reliance on traditional methods.

Sugarcane is a prominent cash crop in Pakistan, ranking among the top five. It contributes approximately 0.6% to the country's GDP [2]. During the 2020-2021 period, sugarcane cultivation spanned an area of 1.2 million hectares, resulting in a production of 80 million tons as presented in Table 1.1. Apart from its primary use in sugar and jaggery production, sugarcane offers other valuable byproducts such as ethanol, alcohol, bagasse, and press mud. Ethanol is utilized as fuel, while alcohol finds applications in the pharmaceutical industry. Bagasse, the fibrous residue, serves as a source of renewable energy for electricity generation, and press mud is used to enhance soil fertility.

**Table 1.1: Sugarcane Area and Production**

Sr	Year	Area		Production		Yield	
		(000 Hectare)	% Change	(000 Tons)	% Change	(Kgs/Hec.)	% Change
1	2016-17	1218	-	75482	-	61972	-
2	2017-18	1342	10.2	83333	10.4	62096	0.2
3	2018-19	1102	-17.9	67174	19.4	60956	-1.8
4	2019-20	1040	-5.6	66380	-1.2	63827	4.7
5	2020-21	1165	12.0	81009	22	69536	8.9

In Pakistan, the production of sugarcane is distributed across various regions, with the Punjab province contributing the majority at 66 percent, followed by Sindh contributing 26 percent, Khyber Pakhtunkhwa (KPK) contributing 8 percent, and Baluchistan with almost 1 percent as demonstrated in Table 1.2. Additionally, in the higher elevations of KPK, due to the more temperate climate there is a low production of sugar beet. Farmers in Pakistan commonly prefer to plant sugarcane

either in the autumn or spring, autumn planting typically yielding superior outcomes due to the extended growing season.

Sugarcane cultivation practices vary in different regions of Pakistan. In Punjab and KPK, sugarcane is planted during the spring season and harvested it after a duration of 8-10 months. Conversely, in Sindh, the majority of planting occurs during the autumn, providing an extended growth period of up to 16 months. This longer growth period in Sindh contributes to a slight enhancement in the sucrose content of the sugarcane, potentially resulting in better prices when selling to sugar mills[3].

**Table 2.2: Sugarcane Area and Production with respect to Province**

Sr	Year	Area (Hectare)			Production (Tons)		
		2019/20	2020/21	2021/2022	2019/2020	2020/2021	2021/2022
1	Punjab	670K	784K	845K	44M	51M	57M
2	Sindh	265K	280K	310K	17M	17.9M	19.8M
3	KPK	102K	110K	114K	5.7M	5.8M	6M
4	Baluchistan	1K	1K	1K	0.047M	0.047M	0.048
5	Total	1,038K	1,175K	1,270K	67M	75.5M	83M

### 1.3 Challenges in Sugarcane Production and Management

The productivity of sugarcane per-hectare in Pakistan falls below the international standards. Specialists attribute this to various factors such as water shortages, limited availability of high-yielding varieties, and inconsistent application of fertilizers and pesticides. To improve yields and bring them in line with global standards, its widely agreed that Pakistan would have benefitted impact from increased research and development initiatives focused on sugarcane and crop monitoring and mapping techniques.

Moreover, during the process of field data collection, there is a possibility of errors occurring, which can result in inaccurate and improper measurements. Additionally, in certain areas where physical access is limited, estimation is carried out instead. Unfortunately, these circumstances can lead to the transmission of incorrect information to sugar mills, potentially causing a shortfall in the crushing process. As sugar mills operate based on specific timeframes, any inaccuracies in the transmitted information can disrupt the smooth functioning of the mills, leading to unforeseen challenges in meeting the required crushing quotas.

In order to get accurate crop maps at large scale satellite imagery along with efficient Deep learning algorithms have been showing promising results. This study will not only help sugar mills but also farmer with large crop areas to manage resources and estimate production.

## 1.4 Satellites used in Crop Classification

Various satellites have been launched that is used in crop mapping and management, vegetation health monitoring, crop yield estimation and monitoring soiland irrigation requirements. A comparison of most popular satellites that has been used in satellite enabled farming is shown in Table 1.3.

**Table 3.3:** Satellite Comparison

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

Sentinel-2 has the highest spatial resolution among the satellites whose data is freely available. Additionally, its temporal resolution of 5 days can be regarded as nearly real-time for crop mapping purposes. Numerous studies have utilized Sentinel-2 and have observed encouraging outcomes.

## 1.5 Research Objectives

This study proposes deep learning-based framework for classifying sugarcane crop among other crops grown in Pakistan. To achieve this goal, the study sets out the following objectives:

- Develop a robust framework that is capable of classifying and mapping sugarcane crop with great accuracy among other crops irrespective of area.
- Select sentinel products based on time-series NDVI plot.
- Extract subpixel information for labeling of each pixel using spectral unmixing.



- Develop 2D time-series spectral feature map of each pixel which is further used for classification.
- Explore and implement multiple classification algorithms within the framework.
- Evaluate the performance of the developed frame work using performance metrics

<p>Chapter 1 Introduction</p>	<ul style="list-style-type: none"> <li>• Overview</li> <li>• Contribution of Sugarcane to Pakistan's Economy</li> <li>• Challenges in Sugarcane Production and Management</li> <li>• Satellites used For Crop Mapping</li> <li>• Research Objectives</li> </ul>
<p>Chapter 2 Literature Review</p>	<ul style="list-style-type: none"> <li>• Crop mapping using conventional ML techniques</li> <li>• Crop mapping using time weighted dynamic time warping method</li> <li>• Crop mapping using DL techniques</li> <li>• GAP Analysis</li> </ul>
<p>Chapter 3 Study Area and Materials</p>	<ul style="list-style-type: none"> <li>• Study Area</li> <li>• Spectral Indices</li> <li>• Spectral Unmixing</li> <li>• Classification model</li> </ul>
<p>Chapter 4 Proposed Framework</p>	<ul style="list-style-type: none"> <li>• Preprocessing</li> <li>• Feature Extraction</li> <li>• Labeling</li> <li>• Classification</li> </ul>
<p>Chapter 5 Results &amp; Discussion</p>	<ul style="list-style-type: none"> <li>• Evaluation metrics</li> <li>• Classification maps for dataset combination 1</li> <li>• Classification maps for dataset combination 2</li> </ul>
<p>Chapter 6 Conclusion</p>	<ul style="list-style-type: none"> <li>• Shortcomings Identification</li> <li>• Proposed solutions as future work</li> </ul>

**Figure 1.1: Flowchart of the Research**

# Chapter 2: Literature Review

This chapter describes the extensive review of the research work done to facilitate the agriculture sector in adapting modern and recent techniques, among which crop mapping and monitoring are the most essential. The launch of Landsat-1 opened new gates for the research community to utilize the satellite data to get crop maps at very low cost and at very large scale.

This chapter is sectioned in three parts, first section presents the work done using conventional machine learning techniques and second part defines the work related to time weighted time warping method to obtain the classification maps and final part describes the work done in deep learning domain.

## 2.1 Crop mapping using conventional machine learning techniques

Previous studies have demonstrated the efficacy of traditional machine learning algorithms, including Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbour (KNN), and Decision Tree (DT), in crop classification tasks, yielding favorable results compared to traditional methods. However, these approaches require substantial effort in terms of domain knowledge and feature engineering to extract relevant information from raw data.

### 2.1.1 Incorporating single-date data

Soon after the launch of sentinel 2, pixel based and object-based classification of crops and tree species were carried out using random forest algorithm utilizing single date images. The crops having greater training data seemed to be classified with better accuracy and pixel-based classification was slightly more accurate than object-based classification. Study also stated that red-edge and shortwave infra-red bands contributed the most towards better crop mapping [4].

The classification of six different crops were performed in the western region of Japan using most widely used machine learning algorithms such as random forest (RF) and support vector machine using multi spectral instrument.

Eighty-two vegetation indices were calculated using MSI to classify the crops with greater accuracy. Along with RM and SVM, ensemble learning technique was used that resulted in more accurate results than the two previously applied algorithms. Moreover, various algorithms were applied to evaluate the accuracy of each method. Since study shows that many researchers have used random forest classifier, because RF is ensemble learning technique that results in good approximation and classification [5].

### **2.1.2 Incorporating time-series data**

Similarly, Performance of several machine learning algorithm was accessed on four types of datasets obtained from sentinel-2, i.e., first being the uni-temporal images, 2<sup>nd</sup> contained combination of 5 best performing images in single date, 3<sup>rd</sup> comprises of handpicked 5 images that covered the crop developing period, and 4<sup>th</sup> contained images covering the chronological stage before harvesting. In order to train the classifiers forty one features were generated from single sentinel image and accuracies were calculated. In almost all of experiments SVM outperformed the rest [6].

### **2.1.3 Incorporating fusion of different satellite data**

This study, examines the potential of integrating Sentinel-1 SAR data with Sentinel-2A optical data to enhance the efficiency of crop classification. They assess the importance of different input data and investigate the temporal dynamics of various crop types. Nine scenarios are tested, including both SAR-only and SAR-optical integration approaches. The study evaluates the effectiveness of 22 nonparametric classifiers is assessed, focusing on previously untested algorithms with SAR data. The findings suggest that the most effective scenario involves integrating VH and VV SAR data with NDVI and employing a cubic support vector machine (SVM) classifier. This combination yields the highest accuracy compared to other tested scenarios [7].

To tackle the challenge of crop classification in extensive areas, a novel approach is proposed that combines multi-temporal remote sensing data from open-source platforms such as Sentinel-2 (S2) and Landsat-8 (L8) satellites. A unique concept of Tuplekeys is introduced, enabling the integration of data from these sensors with varying spatial and temporal resolutions.

The fluctuations in the time-series Normalized Difference Vegetation Index (NDVI) serve as input for crop classification. Additionally, filters based on farm management and composition criteria are employed to enhance efficiency and minimize redundancies.

Three machine learning based classifiers are assessed, utilizing a field-based calibration and a pixel-based approach for the ultimate classification. The proposed methodology is tested on the data taken during spring and summer of 2017 from the Duero River basin. Among the analyzed classifiers, the Ensemble Bagged Trees (EBT) algorithm achieves the highest overall accuracy, with individual crop classification at 87% and grouped classification at 92% [8].

A parcel-based approach was employed to assess the performance of 22 nonparametric classification algorithms in Spain. The mean normalized difference vegetation index (NDVI) along with standard deviation were utilized for each plot. Ground truth data from over two thousand visited fields encompassing 12 different crops were used. The study revealed that ensemble classifiers demonstrated robustness but were less efficient, whereas nearest neighbor methods and support vector machines struck a better balance between robustness and efficiency. The overall F1 score, indicating accuracy, approached 90%, although misclassifications were observed for spring crops. Nonetheless, the developed tool successfully distinguished crops with similar growth cycles, such as distinguishing between purple garlic and white garlic [9].

## **2.2 Crop Mapping using time weighted dynamic time warping method**

The application of a time-weighted dynamic time warping (TWDTW) method is explored in both spatial-based and patch-based classifications of diverse crop categories across three discrete study areas of USA, Romania and Italy. The classification results of TWDTW are compared with those of Random Forest (RF) for both spatial-based and patch-based analyses. Patch-based TWDTW surpasses spatial-based TWDTW in all investigated areas, demonstrating overall accuracies ranging from 78.05% to 96.19% and exhibiting faster computational time. In Romania and Italy, TWDTW performs similarly to RF, while RF achieves superior results in the USA due to significant in-class spectral variations. In Addition to,

TWDTW proves to be less dependent on training samples, offering an advantage in regions with limited availability of training data [10].

Temporal patterns were developed by utilizing four parameters extracted from Sentinel-1 SAR data and correlating them with unclassified satellite imagery using the time-weighted dynamic time warping (TWDTW) algorithm. The study focused on pixel-based and parcel-based classifications to identify four different crop varieties under two water limiting conditions. The overall accuracy achieved for pixel-based and parcel-based classifications was 63% and 76% respectively, with corresponding Kappa coefficients of 0.58 and 0.73. The findings suggest that the parcel-based TWDTW algorithm, influenced by temporal patterns of the radar vegetation index (RVI), effectively delineates croplands subjected to varying irrigation treatments. This makes it well-suited for modeling studies related to yield assessment and damage analysis [11].

Another research endeavor focuses on delineating the geographical extent of sugarcane cultivation areas in China utilizing a phenology-based methodology. This approach leverages the temporal patterns observed in time-series data derived from Landsat and Sentinel-1/2 images, which are accessed through the Google EarthEngine platform. By examining the phenological resemblance within the normalized difference vegetation index (NDVI) sequences, the method effectively discerns the presence of sugarcane plantations. The time-weighted dynamic time warping (TWDTW) technique is employed to account for the variable life cycles exhibited by sugarcane crops. The outcomes of this investigation exhibit notable levels of overall accuracy, successfully mapping the distribution of sugarcane areas in concordance with agricultural statistical zones [12].

### **2.3 Crop mapping using deep learning techniques**

Over the decade we noted a shift from extracting features to developing architectures, as ML involves a lot of data engineering and data cleaning where as deep learning algorithm has the capability of extracting features on its own. For a long time, convolutional neural network ruled in solving the computer vision's domain problem. In the same way recurrent neural networks and LSTM was considered as saviors in dealing with sequence data and natural language processing tasks.

### 2.3.1 Using Recurrent Neural Network and LSTM

The correlation within the timeseries data is extracted using recurrent neural networks, while features are extracted using convolutional neural networks to enable classification. By combining these two networks, specifically the recurrent and convolutional neural networks, various crops were successfully classified using a fusion approach.

An innovative deep learning model was introduced that combines Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) specifically designed for pixel-based land cover and crop classification. The model leverages multi-temporal Sentinel-2 imagery captured in the central north part of Italy, characterized by diverse agricultural systems. By automatically extracting features through learning time correlations, the proposed methodology reduces the reliance on manual feature engineering. Impressively, the model achieves an outstanding overall accuracy of 96.5%, surpassing the performance of traditional ML algorithms like SVM, RF, Kernel SVM, and XGBoost [13].

Various deep learning architectures were examined using time-series data extracted from Multi-temporal Sentinel-2 imagery. The assessed architectures encompass

- one-dimensional convolutional neural networks (1D-CNNs),
- two-dimensional CNNs (2D-CNNs),
- three-dimensional CNNs (3D-CNNs),
- long short-term memory (LSTM), and
- two-dimensional convolutional LSTM (ConvLSTM2D).

The findings illustrate that 1D-CNN and LSTM models outperform random forest (with accuracies of 92.5% and 93.25%, respectively) when utilizing a single time-based attribute. The 2D-CNN model, incorporating both temporal and spatial information, achieves a little higher accuracy of 94.76%, but it falls short in fully leveraging the multi-spectral features. Integrating both temporal and multi-spectral features enhance accuracy to 96.94% for the 1D-CNN model and 96.84% for the LSTM model. However, both models do not sufficiently extract spatial information. On the other hand, the 3D-CNN and ConvLSTM2D models achieve higher accuracy rates of 97.43% and 97.25%, respectively, by effectively utilizing both temporal and spatial information [14].



### **2.3.2 Using Convolutional Neural Network**

The fusion of Sentinel-2 optical and Sentinel-1 Synthetic Aperture Radar (SAR) imagery was employed to accurately detect the distribution of paddy rice fields. A workflow was created using CNN based U-Net segmentation and the GEE geospatial analysis platform to delineate smallholder paddy rice fields. The most suitable dataset for mapping paddy rice was determined to be a fusion of Sentinel-2 multispectral bands and Sentinel-1 SAR dual polarization bands acquired during the crop growing season. The deep learning based segmentation model, ResU-Net, achieved an overall accuracy of 94%, surpassing the performance of the random forest (RF) classification method. Time series images were created by extracting pixel-level data from Sentinel-2 satellite images and processing all available bands over multiple dates to classify crops with greater accuracy. A deep convolutional network system is trained with historical data to extract the essential features for classifying distinct crop varieties throughout the year. The proposed method allows for efficient crop classification while being computationally low-cost [15].

A novel framework for pixel classification was introduced, utilizing spectral and time-series data obtained from the Sentinel-2 satellite to map two rice varieties, Basmati and IRRI, cultivated in Pakistan. Data utilized in this study was rice fields at various time points throughout the entire rice-growing season. To extract spatial information for labeling each individual pixel a linear spectral unmixing model was incorporated. The classifier was fed a 16x15 image containing spectral features from different time points and generated pixel-level classifications for Basmati, IRRI, as well as other classes such as soil and water. Experimental findings showcased an impressive overall accuracy of 98.6% for the anticipated methodology, with Basmati rice achieving a higher accuracy of 99.7% compared to IRRI rice, which achieved 95.2% accuracy [16].

### **2.3.3 Convolution Neural Network with Attention Layers**

The researchers examine and contrast three integration approaches (input, layer, and decision levels) to identify the optimal method for enhancing the classification performance of combined optical-radar data. They employ the pixel-set encoder-

temporal attention encoder (PSE-TAE), a recently established architecture designed for parcel-based classification of spatio-temporal satellite imagery using self-attention mechanisms. The experiments are conducted in Brittany, France, utilizing temporal data from Sentinel-1 and Sentinel-2 satellites. The findings demonstrate that both the input and layer-level fusion strategies surpass the decision-level fusion, achieving the highest overall F-score with a notable 2% improvement. Decision-level fusion enhances the accuracy results of prominent classes, while layer-level fusion significantly enhances the accuracy of less prominent classes, exhibiting an enhancement of up to 13%. In comparison to using data from a single sensor, the fusion strategies yield more precise identification of crop types [17].

A new attention-based CNN methodology named Geo-CBAM-CNN was developed to achieve precise crop classification utilizing time series imagery from the Sentinel-2 satellite. This approach was designed to tackle the challenges posed by geographical variations and a vast number of features. By integrating geographic information into an advanced attention module called CBAM, the Geo-CBAM module effectively mitigated the impact of heterogeneity and filtered out unnecessary information. The proposed model surpassed three other cutting-edge approaches, exhibiting outstanding performance with an overall accuracy of 97.82%, a Kappa coefficient of 96.82%, and a Macro-average F1 score of 96.96%. It showcased remarkable spatial adaptability and assigned attention weights to different features, with a higher emphasis on red-edge features. The Geo-CBAM-CNN model exhibits significant potential for large-scale crop mapping applications [18].

## **2.4 Research Gap Analysis**

Extensive research has been carried out using Sentinel imagery to facilitate crop mapping and classification through the utilization of machine learning (ML) techniques, time-warp methods, and various deep learning algorithms. ML techniques often involve extensive feature extraction and engineering. However, over the past decade, a shift has been observed towards the development of deep learning architectures that require less preprocessing or feature extraction compared to ML. CNNs have been predominantly used for image classification, and they can serve as a backbone for various other techniques such as image segmentation, object detection, image generation, and image captioning.

Literature review reveals that CNN, along with other algorithms, has been employed to achieve improved results. For instance, CNN combined with RNNs, Long Short-Term Memory (LSTM), or attention layers, are computationally intensive yet effective for handling complex images. In our proposed framework, we address these challenges by reducing the large images with 13 channels to 2D time-series spectral maps, where each spectral map represents an individual pixel. This approach eliminates the need for merging different deep learning architectures. Additionally, most studies commonly perform training and testing on the same area and a single growing season.

Considering the aforementioned complexities and limitations, we propose a model with the ability to classify crops in different study areas and growing seasons, aiming to evaluate its robustness. Our specific focus lies in the pixel-based classification of sugarcane, utilizing a reduced set of Sentinel products. Our approach combines multispectral patterns and the temporal evolution of each pixel, organizing them into a 2-D image. To capture temporal dependencies, we leverage the features of recurrent neural networks and LSTM, enabling each pixel to retain its previous state information. Alongside spectral bands, we incorporate six biophysical and spectral indices, which are stacked together to enhance the efficiency of the developed model. The resulting images are simplified yet contain sufficient information for straightforward classification using a simple multilayer perceptron or convolutional neural network, particularly for the same district. When dealing with different districts, we explore various options, including state-of-the-art CNN architectures and LSTM. Despite dataset variations such as location and growing season, the classification results prove to be satisfactory. This study contributes to the identification of sugarcane at any stage of its life cycle and facilitates the estimation of sugarcane availability in a specific area using a reduced number of Sentinel products.

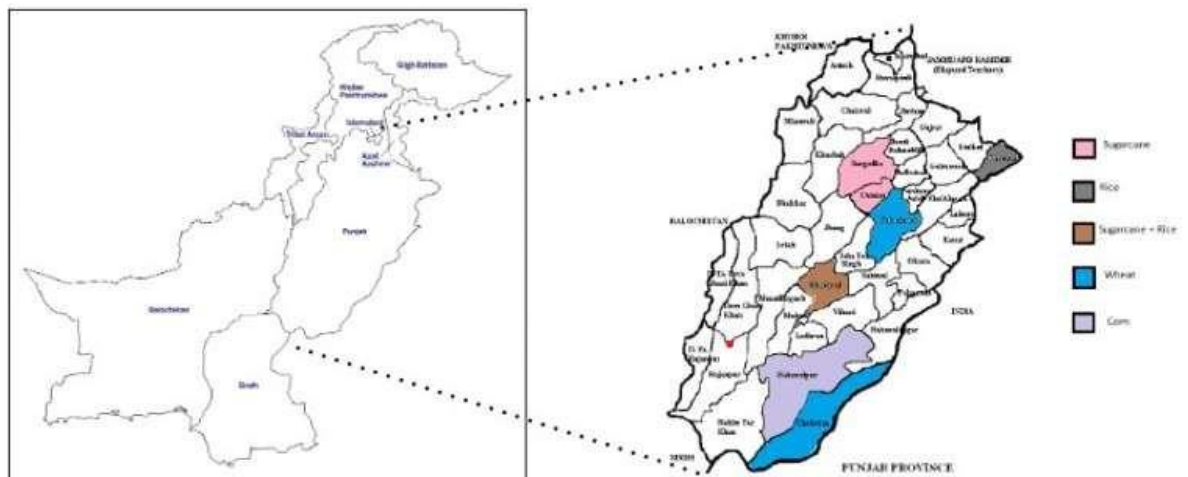
# Chapter 3: Study Area and Materials

This chapter primarily centers around the study area, providing a comprehensive exploration of the geographical region under investigation. It delves into the key concepts and principles that underpin the research conducted in this study. By explaining these fundamental aspects, the chapter aims to provide a solid foundation for understanding the subsequent analyses and findings presented in the following chapters.

## 3.1 Study Area

The study area we focused on encompasses six districts in Punjab, Pakistan, spanning from 29.3544° N, 71.6911° E to 33.5651° N, 73.0169° E. This region experiences a temperature range of 0 to 26 °C during winter and 28 to 48 °C during summer, with an annual precipitation of 532.384 mm.

For our research, we collected field data specifically from the districts of Chiniot, Sargodha, and Khanewal, which provided information about sugarcane cultivation. Additionally, field data for other crops such as rice was gathered from Narowal and Khanewal, wheat from Faisalabad and Cholistan, and corn from Bahawalpur.



**Figure 3.1:** Study area

To support our analysis, we acquired Sentinel-2A products from the open-access hub [19]. These products were selected based on their cloud coverage, with a criterion of less than 8%. Each product covers an area of 100x100 km<sup>2</sup>, and there are two types

available: 1C and 2A. The Sentinel-2, 1C product consists of thirteen spectral bands, with four bands offering a spatial resolution of 10 m, six bands having 20 m, and three bands have 60 m. The Sentinel-2, 2A product includes 12 bands, with band 10 specifically used for atmospheric correction to derive the bottom of atmosphere reflectance. Georeferenced field data and sowing dates were obtained as ground truth from Agriculture Robot Lab at NCRA [20].

The study area consists of a total crop field of 270 acres. Among which sugarcane spans the area of 130 acres whereas wheat, rice and corn fields cover an area of 45, 55, and 42 acres respectively. It is evident from figure 1 that study area is distributed through the Punjab province and not confined to one particular district or area as in most of studies one particular area is selected and trained and evaluated on same region.

### **3.2 Spectral Indices**

Spectral indices play a crucial role in crop mapping and classification by utilizing mathematical ratios of different spectral bands within multispectral imagery. These indices are applied to individual pixels, allowing for the extraction of key features related to various crop types.

The normalized difference vegetation index (NDVI) is a commonly employed spectral index that evaluates the density and health of vegetation. It accomplishes this by quantifying the disparity among the reflectance values of the near-infrared (NIR) and red spectral bands captured in multispectral imagery. By comparing the reflectance of these specific wavelengths, NDVI offers valuable insights into the vigor and abundance of vegetation. This information is particularly useful in assessing vegetation dynamics, monitoring ecosystem health, and detecting changes in land cover over time. NDVI serves as a fundamental tool in remote sensing and plays a vital part in various fields such as agriculture, ecology, and environmental management [21].

Another important spectral index is the normalized difference water index (NDWI), which primarily reflects the presence of water content within vegetation. It calculates the discrepancy between the green and NIR channels and helps to identify drought conditions in crops [22]. By analyzing the spectral variations associated with water content, NDWI serves as an indicator of vegetation stress caused by water scarcity.

The differenced vegetation index (DVI) is a relatively simple spectral index that exhibits sensitivity to changes in vegetation relative to the background, such as soil [23]. By quantifying the discrepancy between reflectance values, DVI enables the detection of vegetation changes and provides valuable information about variations in plant growth and health.

To consider the impact for the soil reflectances on vegetation indices, the soil-adjusted vegetation index (SAVI) includes an adjustment factor to minimize this impact [24]. By mitigating the confounding effects of soil reflectance, SAVI enhances the accuracy of vegetation mapping and monitoring.

In addition to these indices, the leaf area index (LAI) serves as a biophysical indicator that quantifies the ratio of leaf area to ground area. It provides information about the extent and density of leaves within the canopy [25]. LAI is particularly useful for understanding vegetation structure and estimating biomass.

Lastly, the fractional vegetation cover (FVC) is a metric that quantifies the percentage of soil surface covered by green vegetation. It provides an assessment of the spatial extent and distribution of vegetation within a given area.

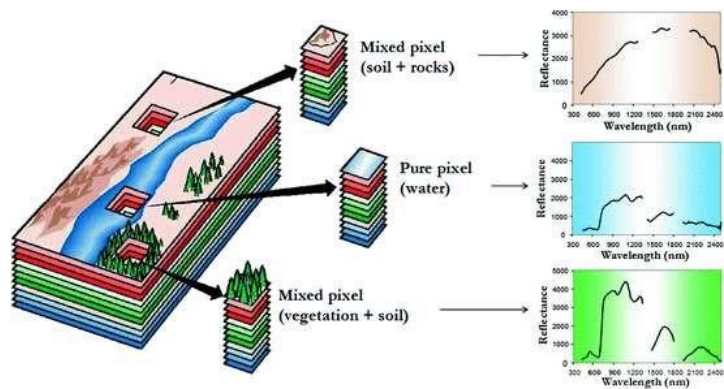
By employing these spectral indices, researchers can harness the unique characteristics of multispectral imagery to accurately map and classify crop types, assess vegetation health, monitor water availability, and derive valuable insights into the dynamics of agricultural ecosystems.

**Table 3.1: Spectral indices ranges**

<b>Spectral Indices</b>	<b>Value range</b>
NDVI	-1 to +1
DVI	-1 to +1
SAVI	-1 to +1
NDWI	-1 to +1
FVC	0 to +1
LAI	1 to 4.5

### **3.3 Spectral Unmixing**

When working with low spatial resolution imagery, a difficulty arises where the pixel size or instantaneous field of view (IFOV) exceeds the size of the object or area of interest. This constraint poses a challenge in accurately discerning and isolating the unique spectral characteristics of individual objects or features. Instead, each pixel's spectral signature becomes a mixture of various materials or features, represented by their individual pure spectral signatures, also known as endmembers (such as water, soil, rock, vegetation, etc.). The fraction of each endmember within a pixel is referred to as fractional abundance.



**Figure 3.2:** Spectral Unmixing[26]

Consequently, each mixed pixel contains a combination of two or more classes, which significantly impacts the accuracy of supervised classification. The presence of mixed pixels introduces complexities and uncertainties in the classification process, as the spectral characteristics of different objects or features are blended within a single pixel. As a result, accurately identifying and classifying specific land cover types becomes challenging, particularly in cases where mixed pixels are prevalent.

Efforts to address this issue involve advanced techniques for sub-pixel analysis, unmixing algorithms, and spectral mixture analysis. These methods aim to estimate the fractional abundances of different endmembers within mixed pixels, allowing for more accurate classification results and improved characterization of land cover types. By effectively handling mixed pixels, the classification accuracy can be enhanced, enabling more precise analysis and interpretation of remotely sensed imagery for various applications in fields like agriculture, land use planning, and environmental monitoring.

### 3.4 Classification Models

### 3.4.1 VGG-16

VGG16 is an architecture of a CNN that comprises of 21 layers, including 13 convolutional layers, 5 Max Pooling layers, and 3 Dense layers. One notable feature of VGG16 is its emphasis on 16 weight layers, which are the parameters that undergo the learning process. The configuration of convolution and max pooling layers in VGG16 follows a consistent pattern, employing 3x3 filters with a stride of 1 for convolutions and 2x2 filters with a stride of 2 for max pooling. The convolution layers are organized with 64, 128, 256, and 512 filters, respectively. After the convolutional stack, three Fully-Connected layers are present, with the first two containing 4096 channels each, and the third performs classification with 1000 channels representing different classes based on the findings discussed in [27]. The architecture concludes with a soft-max layer [28].

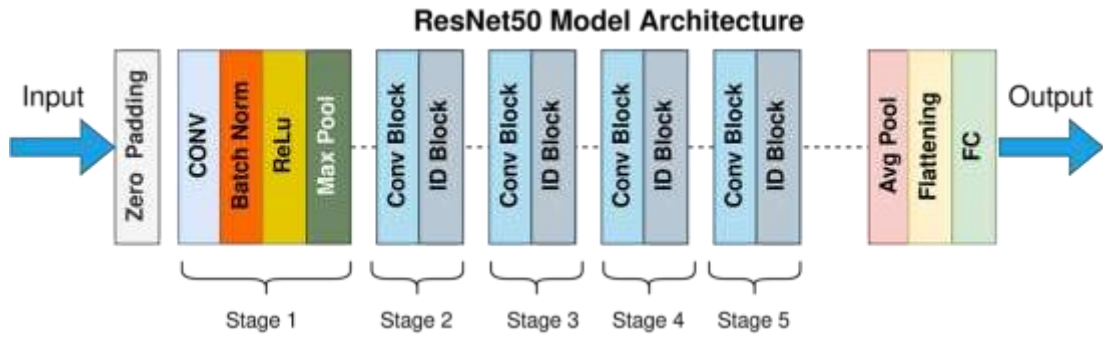


**Figure 3.3:** VGG-16 Architecture [29]

### 3.4.2 ResNet-50

ResNet-50 is an advanced convolutional neural network (CNN) architecture comprising a total of 50 layers. It introduces an innovative approach called residual connections, which effectively tackle the issue of vanishing gradients commonly encountered in deep networks. This technique enables the network to retain and propagate important information throughout the layers, enhancing the overall learning process. ResNet-50 also incorporates a bottleneck design that significantly improves training speed and reduces the number of parameters, leading to more efficient and effective model training [30].

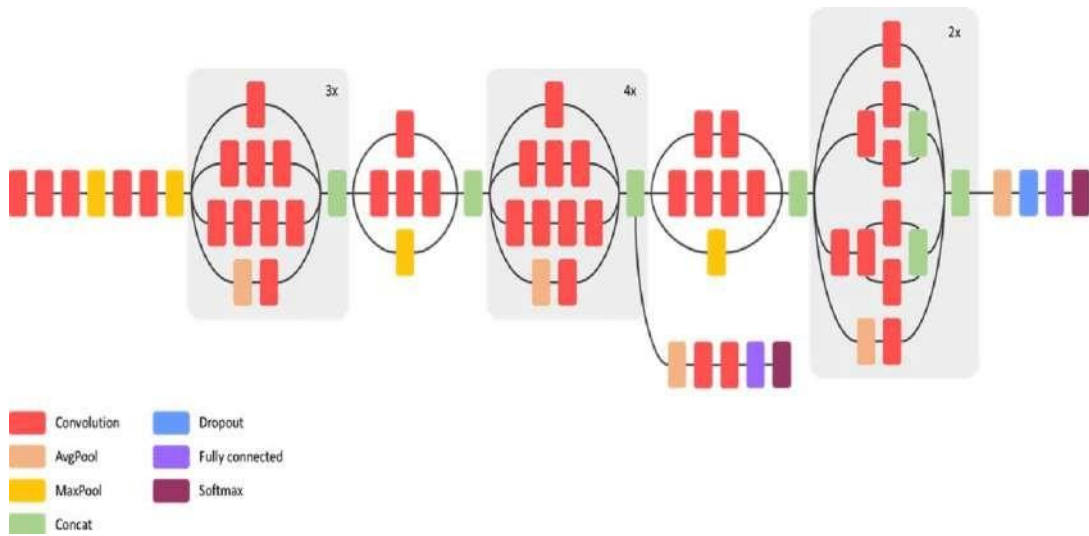




**Figure 3.4:** ResNet-50 Architecture [31]

### 3.4.3 Inception v3

The inception v3 architecture was introduced in the study conducted by [32]. Inception V3 employs inception modules, which consist of parallel convolutional layers with diverse filter sizes, enabling the network to extract features from various spatial scales. Additionally, the architecture incorporates techniques such as factorized convolutions and auxiliary classifiers to enhance training and overall performance.



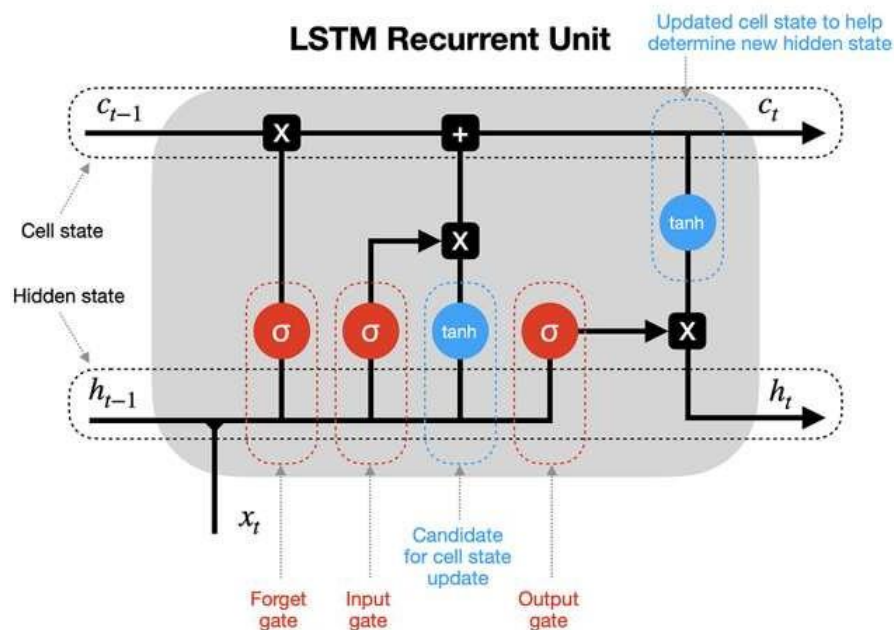
**Figure 3.5:** Inception v3 Architecture[33]

### 3.4.4 Long Short-term Memory

Long Short-Term Memory (LSTM) is a deep learning architecture widely acknowledged for its suitability in handling sequential data and data with extensive dependencies. Unlike traditional feedforward networks, LSTM incorporates feedback connections that enable it to capture context and long-range dependencies within the data. These feedback connections empower LSTM to process entire sequences of

data, rather than treating each data point in isolation. As a result, the model retains information from previous points and combines it with new data points to extract patterns and characteristics. This unique capability makes LSTM a preferred choice for managing sequential and time-series data.

Within LSTM networks, information flow is regulated by gates. These gates play a crucial role in determining which incoming information to accept, what information to retain for future reference, and which information to output from the network. The input gate, forget gate, and output gate collectively govern the flow of information within the LSTM network, ensuring effective management and utilization of relevant data [34].



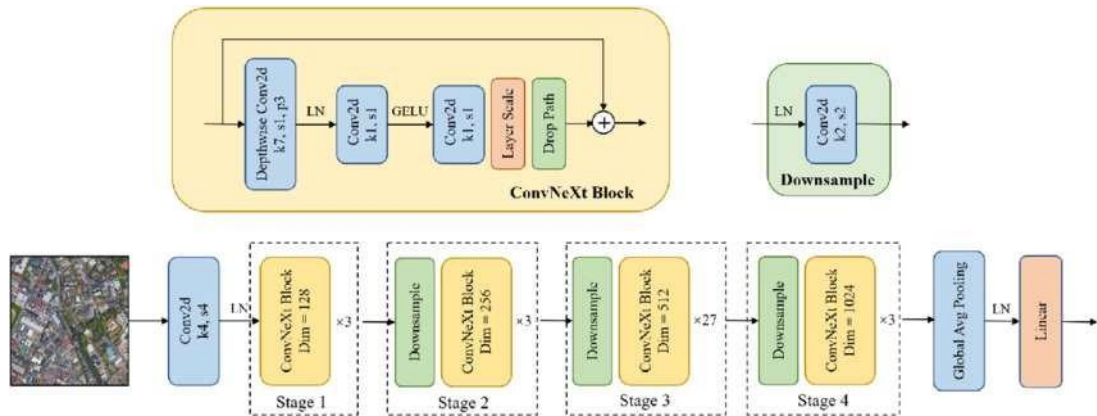
**Figure 3.6:** Long short-term memory Recurrent Unit[35]

### 3.4.5 ConvNext

With the rapid progress of deep learning technologies, traditional recurrent neural networks (RNNs) and Long Short-term Memory (LSTM) networks have largely been replaced by transformers in the field of Natural Language Processing (NLP), as introduced in [36]. Similarly, for more than a decade, convolutional neural networks (CNNs) have been the predominant force in the field of computer vision. Despite the distinct differences between vision and language domains, the architectural principles of these two domains have converged into a new framework called vision transformers, as proposed by [37].

When compared to CNNs in the context of image classification, vision transformers have shown promising results. However, they do not perform as well in image segmentation tasks. To address this limitation, Liu et al. proposed the Swin transformer in 2021, which reintroduced the concept of sliding windows to transformers. The Swin transformer demonstrated strong performance in problems beyond classification, highlighting that the essence of convolutional networks remains significant [38].

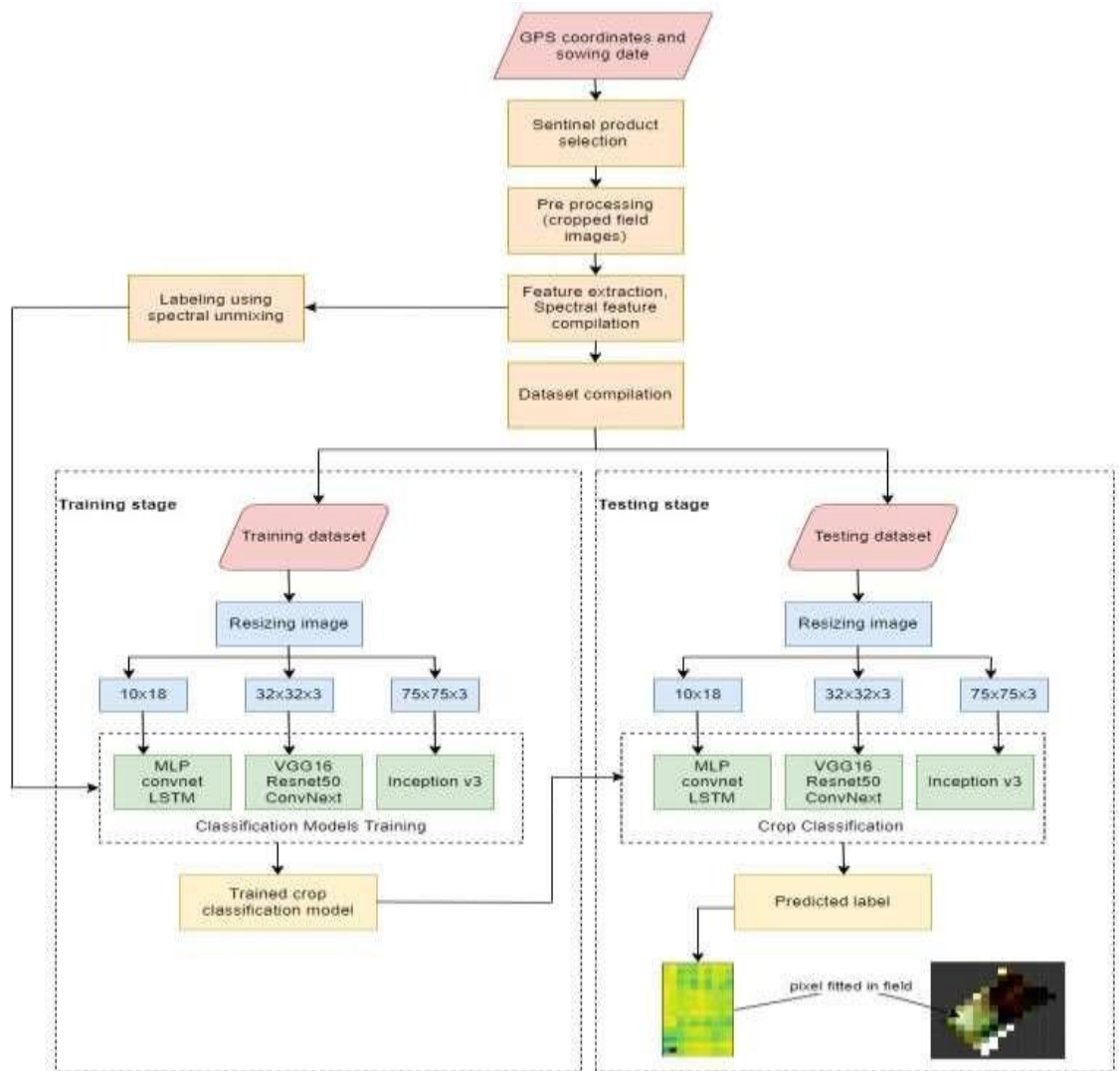
To leverage the advantages of both convolutional networks and transformers, a new architecture called convNeXt was proposed by Liu et al. in 2022. In convNeXt, the baseline architecture is built upon ResNet50, and it progresses with a hierarchical construction similar to the Swin transformer. This approach aims to combine the strengths of both architectures and further enhance performance in various tasks [39].



**Figure 3.7:** ConvNext architecture[40]

# Chapter 4: Proposed Framework

A deep-learning based crop classification framework that is capable of mapping sugarcane crop from other popular crops using a smaller number of Sentinel products is shown in Figure 4.1. and detailed graphical abstract is given in Figure 4.2.

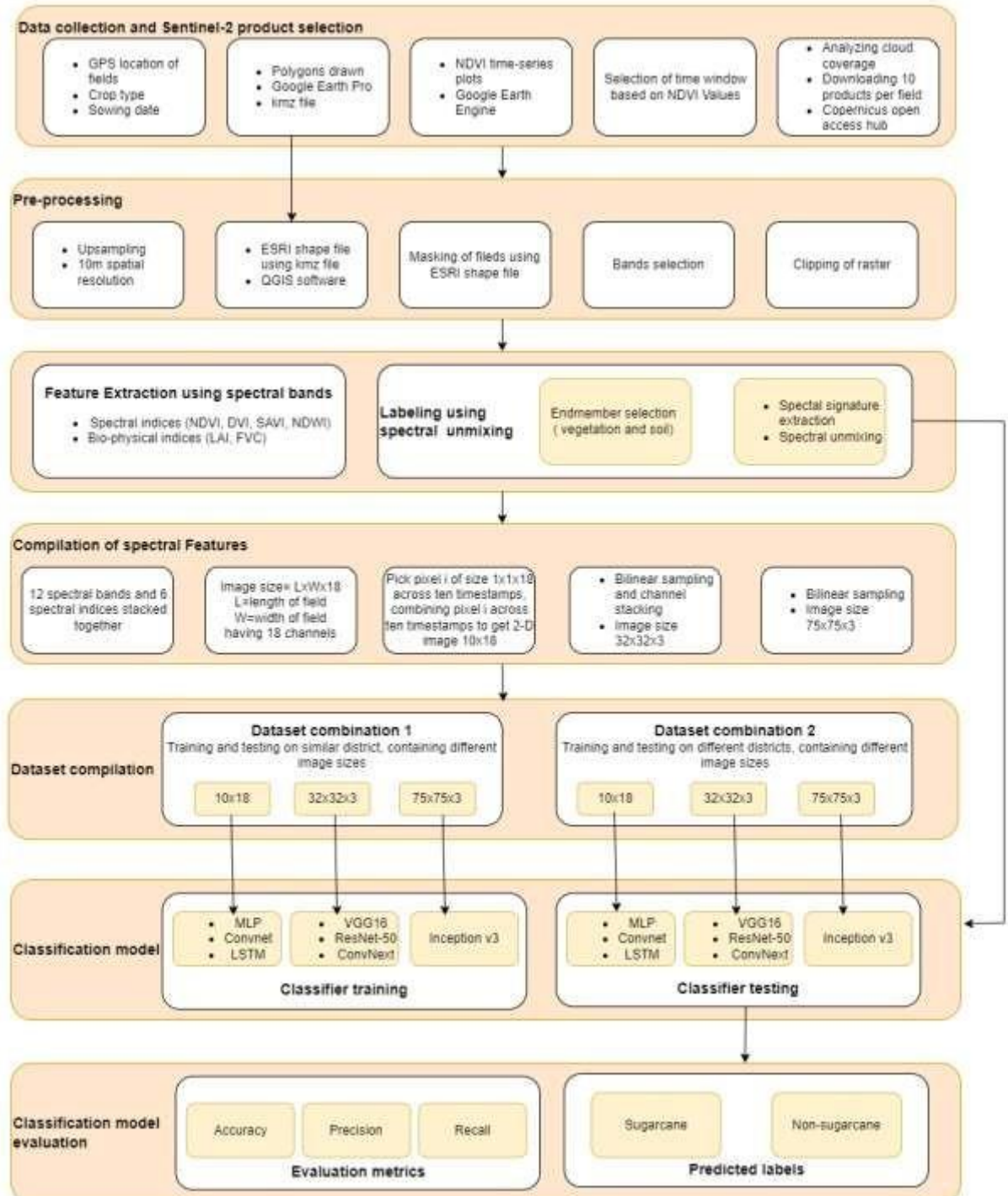


**Figure 4.1:** Block diagram of proposed framework for crop classification

## 4.1 Product Selection

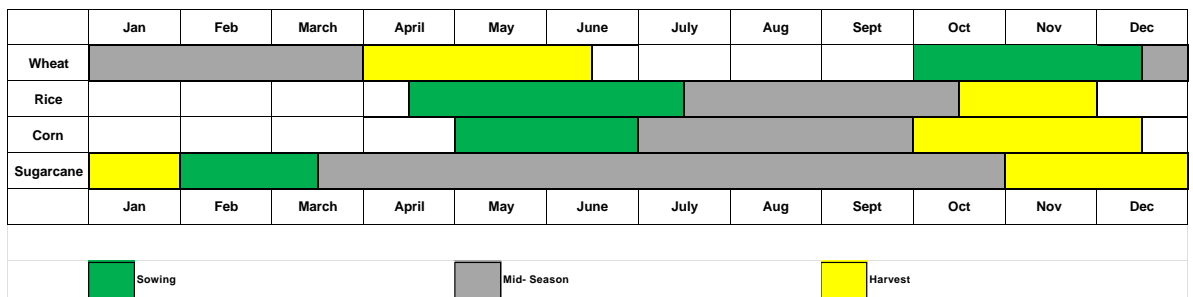
The available data for this study includes georeferenced field data and sowing dates, represented by polygons drawn in Google Earth Pro [41]. NDVI (Normalized Difference Vegetation Index) time series graphs were obtained from the date of

sowing until the estimated harvesting period, as depicted in Figure 4.4. Based on the NDVI crop signatures, a specific time window was determined for each field. Within these time windows, Sentinel products were downloaded, focusing on periods when the NDVI value exceeded 0.2.



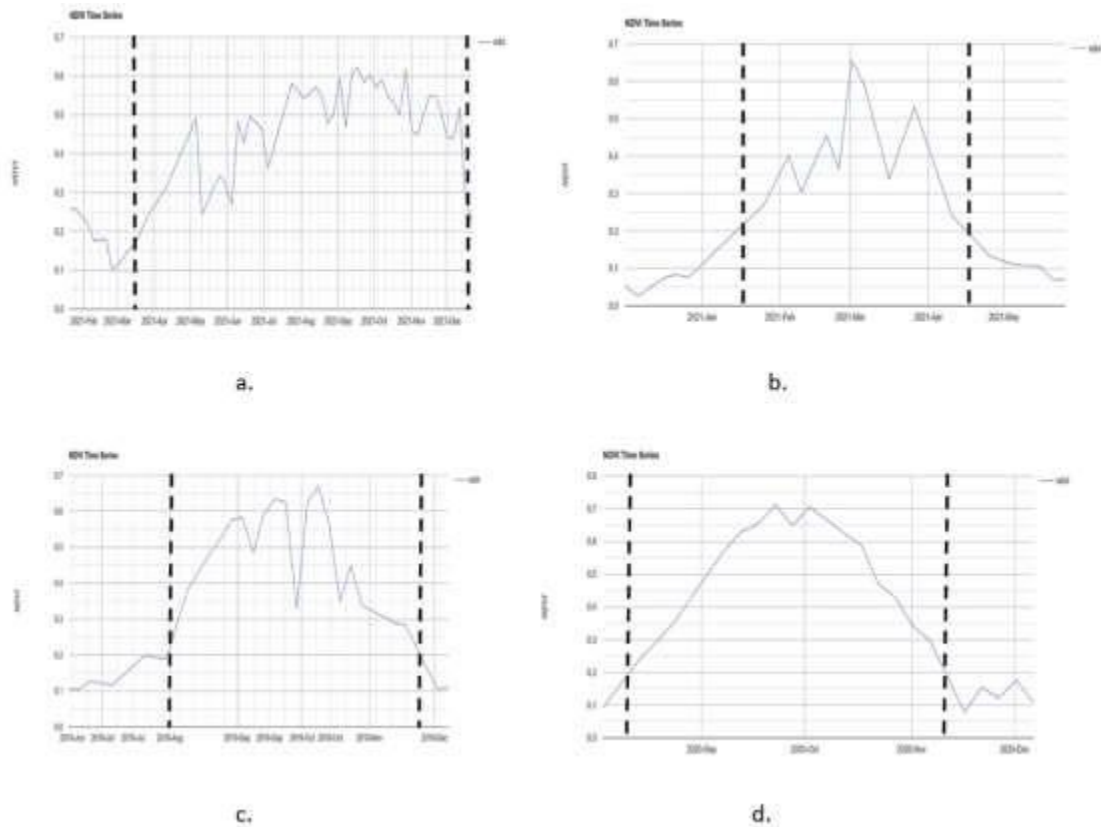
**Figure 4.2:** Detailed Abstract of proposed framework

For each field, a total of ten products were downloaded, covering most of the crop's lifecycle. Since sugarcane in Pakistan typically requires 10 to 12 months to grow, ten tiles with approximately one-month intervals were downloaded. On the other hand, wheat is a 4 to 5-month crop, rice takes around 3.5 to 4 months, and corn has a growth period of 2.5 to 3 months. Consequently, the time duration gap for downloading Sentinel products varied accordingly for each crop type. Figure 4.3 provides a visual representation of the planting and harvesting periods for these four crops.



**Figure 4.3:** Crop calendar of investigated crops

Sentinel products with the cloud coverage of less than eight percent is download and during the month of July and august where cloud coverage is very high in that case the product is analyzed and if the concerned area is cloud free then product is used inthis study.



**Figure 4.4:** Selection of time window for downloading sentinel products for a) sugarcane, b) wheat, c) rice, d) corn

## 4.2 Preprocessing

The pre-processing steps for all Sentinel-2 data were accomplished using the Sentinel Application Platform (SNAP). SNAP is an openly available software tool that provides a modular and feature-rich client platform, portability, efficient memory management, and a framework for graph processing [42]. These features make SNAP an ideal platform for processing and analyzing earth observation data.

**Table 4.1: Crop type and estimated crop life**

Crop type	Crop life in Pakistan
Sugarcane	10-12 months
Rice	3.5-4 months
Wheat	4 – 5 months
Corn	2.5-3 months

The preprocessing and feature extraction steps are conducted utilizing the Sentinel Application Platform (SNAP). To ensure consistency, all bands of Sentinel products with a resolution lower than 10m are up-sampled to a 10m resolution using the bilinear up-sampling method. Reprojection is not necessary since the product is already in the UTM/WGS84 coordinate reference system.

Polygons outlining each field are drawn in Google Earth Pro and then converted to ESRI shapefile format using QGIS software. These shapefiles are subsequently employed to clip the area of interest from the Sentinel products, as depicted in Figure 4.5. Following the preprocessing and polygon clipping procedures, the resulting fields contain 12 spectral bands, including red, blue, green, near-infrared (NIR), short-wave infrared (SWIR), and others.





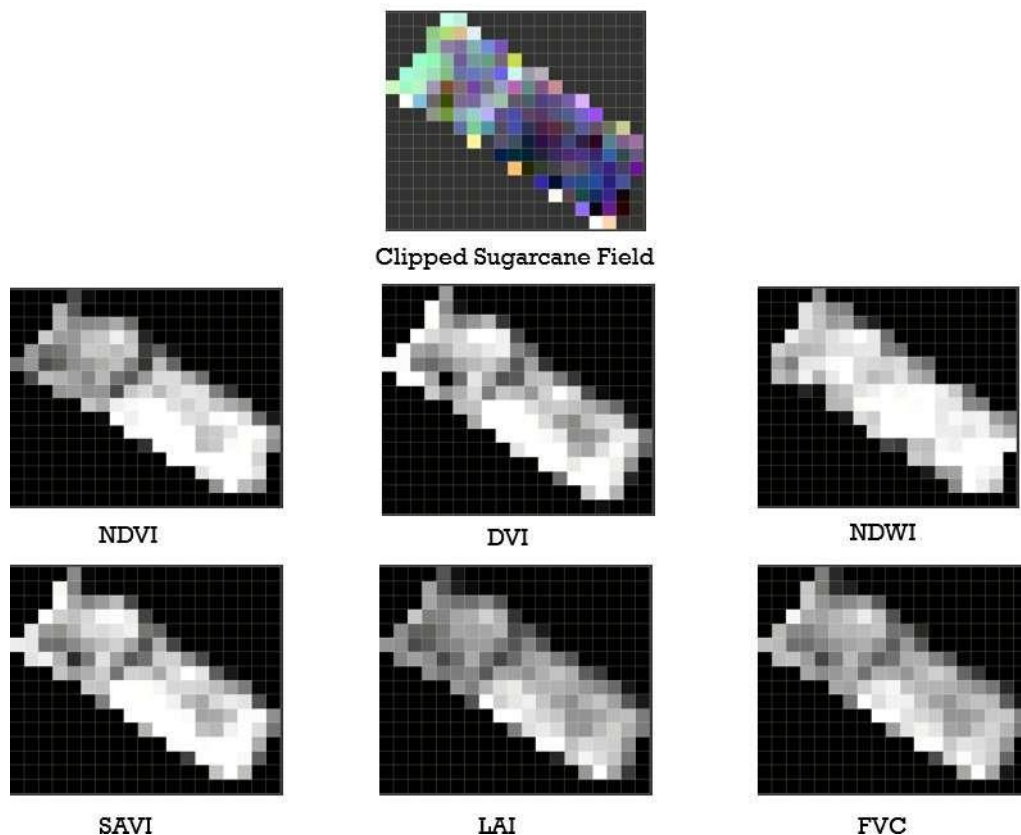
**Figure 4.5:** Polygons used for clipping the fields of a) sugarcane, b) wheat, c) rice, d) corn from sentinel product in SNAP

### 4.3 Feature Extraction

Spectral indices have been used in crop mapping in techniques involving machine learning algorithm. A large number of features are usually extracted, a study mentioned in literature review section extracted around 90 spectral features from the 13 spectral bands of sentinel multispectral imagery. In this study, four spectral indices and 2 bio physical indices, a total of six spectral features along with spectral bands are used. In SNAP, spectral indices such as the

- difference vegetation index (DVI),
- soil adjusted vegetation index (SAVI),
- normalized difference vegetation index (NDVI),
- normalized difference water index (NDWI),
- leaf area index (LAI), and
- fractional vegetation cover (FVC)

are derived from the spectral bands of sentinel product.



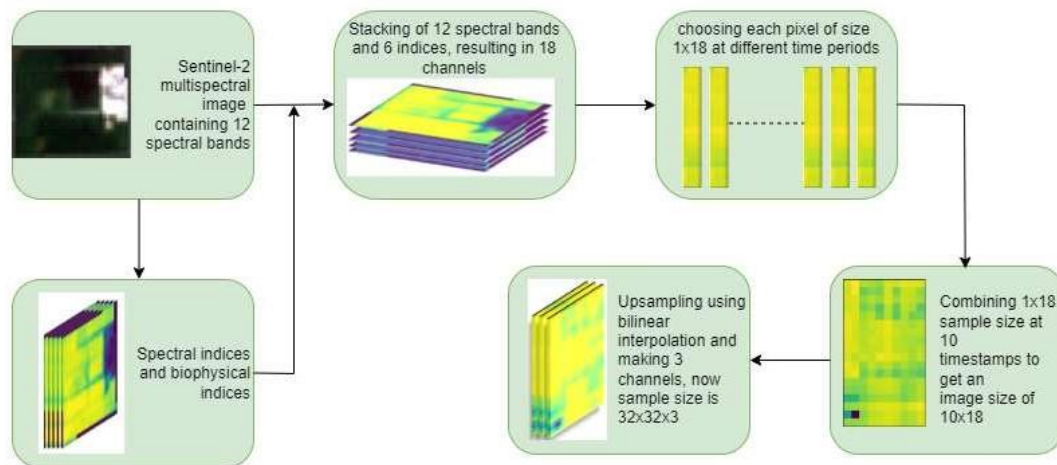
**Figure 4.6:** Spectral features extracted from 12 spectral bands of sentinel image

## 4.4 Compilation of 2D spectral feature maps

The spectral bands and corresponding six indices of the cropped fields are combined by stacking them together. Each pixel is treated independently, where the columns represent the spectral bands and indices, and the rows represent the evolution of that pixel throughout the crop's lifecycle.

To compile the images, a total of 10 Sentinel products are selected based on the NDVI value, covering the entire phenological stages of each crop. These products are combined with the 18 spectral bands and indices, resulting in an image size of  $18 \times 10$  when arranged in a 2D matrix.

For transfer learning purposes, the images are up-sampled using bilinear interpolation, increasing the image size to  $32 \times 32 \times 3$ . This size is suitable for CNN architectures such as VGG16 and ResNet50, as they require a minimum image size of  $32 \times 32 \times 3$ . However, for Inceptionv3, the minimum image size should be  $75 \times 75 \times 3$ . A visual representation of the compilation of spectral features is provided in Figure 4.7, offering a comprehensive graphical overview of the process.



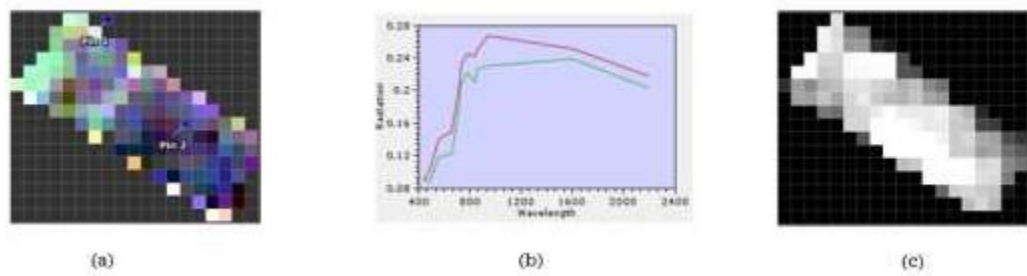
**Figure 4.7:** Compilation of time-series spectral features from field image having 12 spectral bands to  $32 \times 32 \times 3$  image

## 4.5 Labeling

Within multispectral images, pixels that exhibit spectral values resulting from a combination of two or more materials are known as mixed pixels. In contrast, pixels that reflect the spectral signature of a single object are referred to as pure pixels.

To analyze the multispectral images, spectral signatures of green vegetation and soil are extracted using SNAP. Subsequently, an abundance map for each field is

computed at the peak of the crop season. This calculation involves utilizing the two endmembers, namely green vegetation and soil.



**Figure 4.8:** a) Endmember selection, b) Endmember spectrum extraction, c) abundance map of green vegetation

In the green vegetation abundance map, pixel values exceeding 0.5 indicate a high abundance of crop, thereby representing crop pixels. Conversely, pixel values below 0.5 are considered soil pixels, indicating a lower abundance of green vegetation.

#### 4.6 Dataset combination

The dataset used in our study showcases notable disparities in terms of location, sowing year, and crop stage during the selection of Sentinel images. It is a balanced dataset, comprising an approximately equal number of synthetic images for both sugarcane and non-sugarcane classes, which include wheat, rice, and corn.

In the first set of experiments, datasets from all locations are combined and separated into training and validation datasets using a 70:30 ratio. This ensures a representative distribution of data for training and evaluation purposes.

In the second set of experiments, the sugarcane data from Khanewal and Sargodha districts are utilized for training, while the data from Chiniot district is used for testing. Similarly, for other classes, training data is obtained from different districts and tested in different districts. This approach allows us to assess the model's performance across various geographical areas and generalize its effectiveness beyond specific regions.

#### 4.7 Training Classification Models

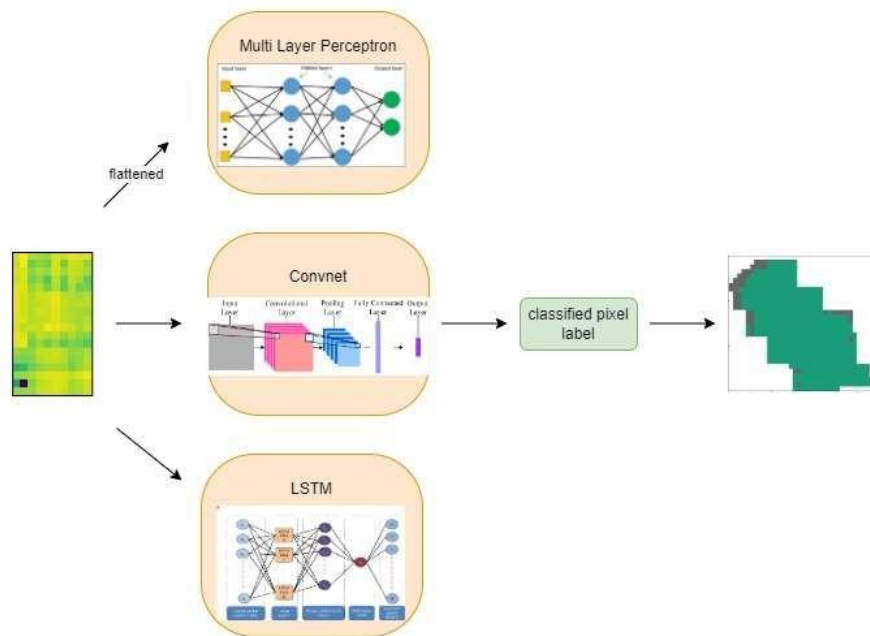
Once the dataset is compiled and divided into training and validation sets, it is subjected to various pre-existing convolutional neural network (CNN) architectures,

including simple sequential convnet, multi-layer perceptron, and LSTM. The CNN models employed in this study encompass VGG16, ResNet50, Inception V3, and convNext.

To leverage the image classification knowledge acquired by these models, the pre-trained portions of the networks are utilized. Specifically, the layers responsible for feature extraction are frozen, while only the classification layers are trained. This transfer learning approach allows the models to classify sugarcane among other crops, benefiting from the previously learned representations.

In the second scenario, all layers of these CNN models are trained using our dataset. This entails updating the weights and parameters throughout the entire network to adapt to the precise features of our dataset and improve the classification performance.

To classify the developed spectral maps with great accuracy several classification algorithms were trained and analyzed which includes Multilayer perceptron, Convnet, state of the art CNN architectures and LSTM. The input size of classifier varies depends on the its architecture. For Multilayer perceptron 10x18 feature map is flattened and fed to the input layer of size 1x180, followed by 3 dense layers and finally a softmax layer.



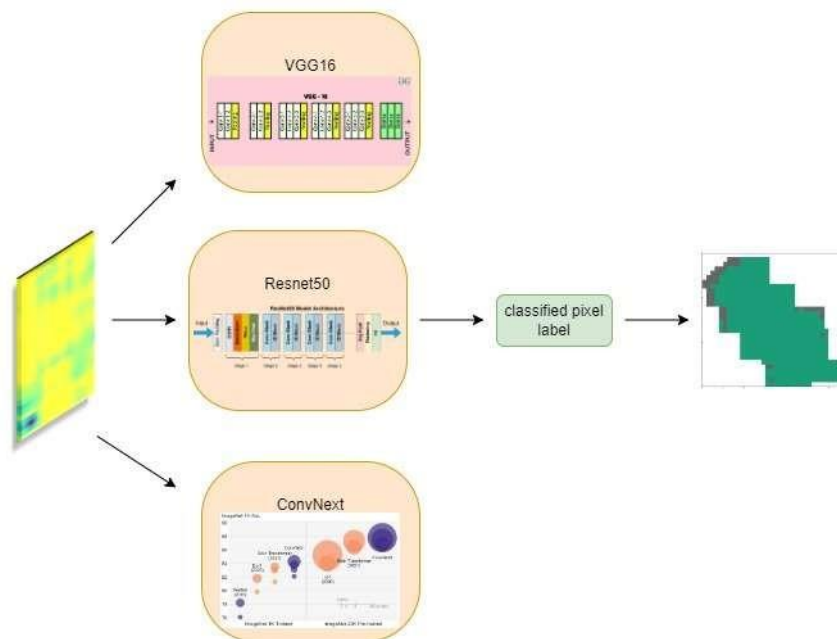
**Figure 4.9:** Classification models trained on sample size 10x18

For convnet the input sample size is a 2D 10x18x1 feature map, the architecture has 3 convolution layers with filter size 32,64,128 and kernel size 3x3 and stride rate of 1. Moreover, there are 2 pooling layers, batch normalization layer, dropout layer, fully connected layer and a softmax layer. The rectified linear unit is served as activation function.

For LSTM input sample size is 10x18, representing 18 spectral features at ten-time stamps. The data was processed using three LSTM layers, with each layer consisting of 20 blocks of LSTM cells and a dropout layer of 0.4. Following the LSTM layers, a linear SoftMax layer with two units, representing the possible number of outcomes to be predicted, is employed.

During the experiments, a learning rate of 0.001 and a batch size of 32 is used consistently. During the training process, the cross-entropy was used as loss function and optimized with an Adam optimizer. The training process spanned 10 epochs, enabling the model to iteratively learn and refine its predictions.

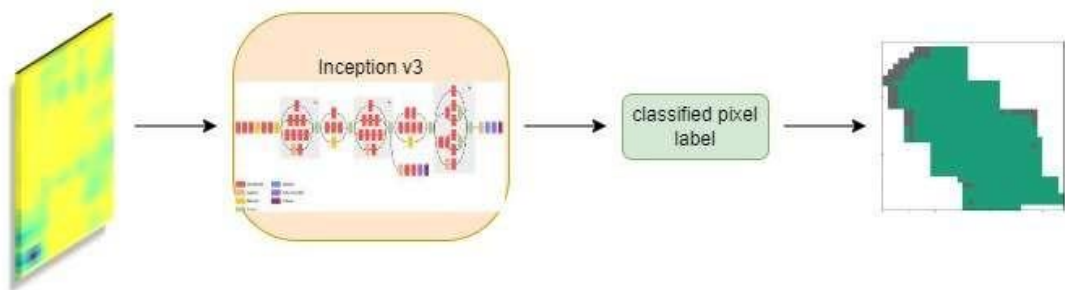
CNN architectures like VGG16, Resnet50, and convNext are trained and optimized using keras framework, the minimum input sample size for above mentioned architectures is 32x32x3 and for Inception v3 its 75x75x3. So, the time-series spectral feature maps are up sampled using bilinear interpolation method to 32x32x3 and 75x75x3. All models were trained using cross-entropy loss function and optimized with Adam optimizer.



**Figure 4.10:** Classification models trained on sample size 32x32x3

To leverage the image classification knowledge acquired by the pre-trained models, that knowledge is utilized for the classification of sugarcane and other crops. In our approach, specifically the pre-trained feature extractor portion of these models is employed, keeping those layers frozen while training only the classification layers.

By utilizing the pre-trained feature extractor, we benefit from the learned representations and patterns extracted from a large-scale dataset. This allows to focus our training efforts on fine-tuning the classification layers to suit the specific task of crop classification. By freezing the feature extraction layers, the valuable knowledge already captured by those layers is preserved and concentrate our training on the final classification stage.



**Figure 4.11:** Classification models trained on sample size 75x75x3

# Chapter 5: Results & Discussion

## 5.1 Proposed Framework for Sugarcane classification

To evaluate the performance of proposed framework for sugarcane crop classification, all the classification models used in this study are tested and analyzed using key evaluation metrics i.e., Accuracy, Precision, and Recall, F1 score. These metrics are represented by mathematical equations, namely Eq. 1, Eq. 2, Eq.3, and Eq. 4, respectively.

### 5.1.1 Confusion matrix

Confusion matrix helps in visualizing the results of classification and provides information about correct and incorrect predictions by the classifier and further used to evaluate the model performance using metrics like accuracy, precision, and recall.

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

**Figure 5.1:** Confusion matrix[43]

The terminology associated with the confusion metrics are described as:

**True positive:** when the actual label corresponds to positive class and is classified as positive class

**False Negative:** when the actual label corresponds to positive class and is classified as negative class

**False Positive:** when the actual label corresponds to negative class and is classified as positive class

**True Negative:** when the actual label corresponds to negative class and is classified as negative class

#### 5.1.1.1 Accuracy

Accuracy is the widely employed evaluation metric in classification, representing the proportion of accurately predicted labels in relation to the total number of labels.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

#### 5.1.1.2 Precision

Precision quantifies the ratio of true positives (TP) to the sum of true positives and false positives (TP + FP). It serves as a measure of the model's accuracy in predicting positive instances, reflecting the ability to minimize false positive classifications and ensure precise positive predictions.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

#### 5.1.1.3 Recall

On the contrary, recall evaluates the ratio of true positives (TP) to the sum of true positives and false negatives (TP + FN). A higher recall value signifies that the model correctly identifies a larger proportion of actual positive cases. It measures the model's capability to effectively capture a significant number of positive instances.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

#### 5.1.1.4 F1 Score

The F1 score is a metric used to evaluate the performance of a classification model. It combines precision and recall into a single value, providing a balanced measure of the model's accuracy. It finds particular application in scenarios where there exists a disparity in the quantity of positive and negative instances within the dataset.

$$F1 \text{ score} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$



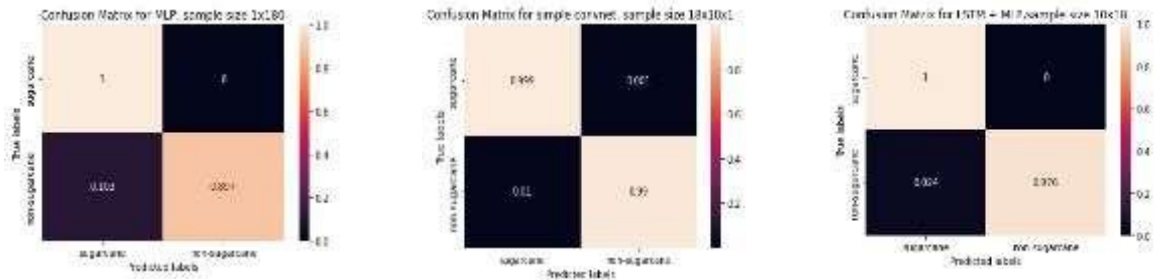
### 5.1.2 Classification results for dataset combination 1

In the initial dataset combination, where the testing data is from one of the districts used for training the classifier, the results are highly impressive. The classification of the 2D time series spectral maps achieves exceptional accuracy. The evaluation metrics of all the models for dataset combination 1 used in the proposed framework are presented in Table 5.1, demonstrating their performance.

**Table 5.1: Evaluation metrics along with input sample size for dataset combination 1**

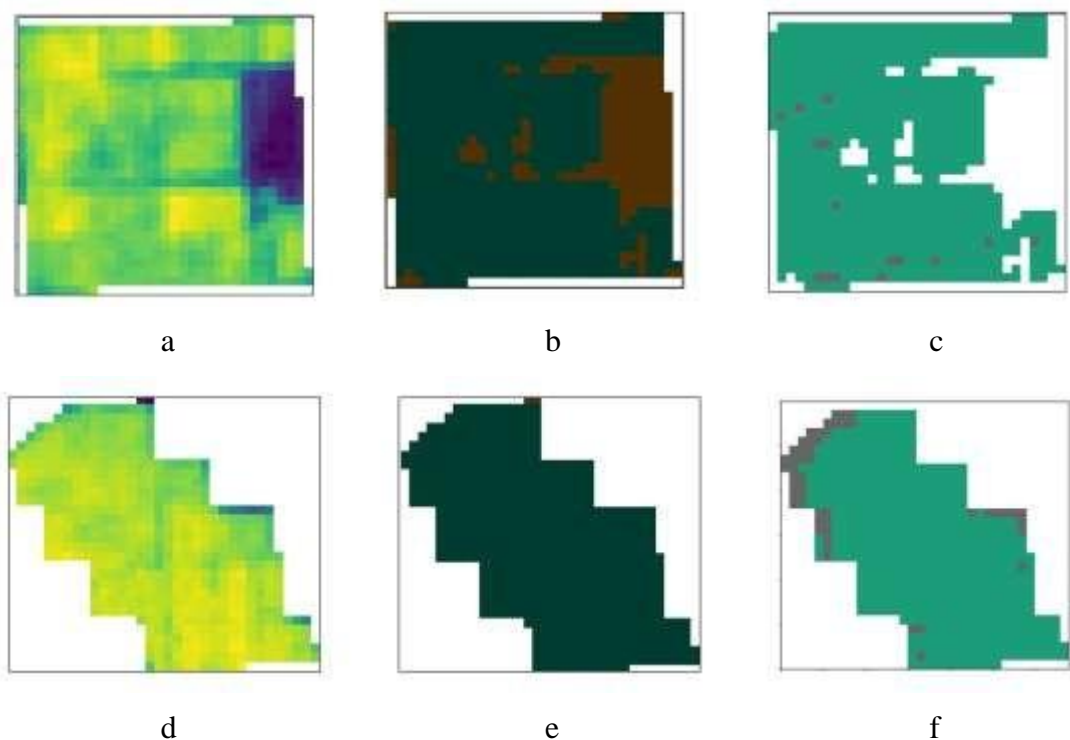
Input sample size	Architecture	Test Accuracy %	Precision	Recall	F1 Score
1x18	MLP	0.77	0.97	0.56	0.71
1x180	MLP	0.94	1.00	0.89	0.94
10x18x1	Simple convnet	0.99	0.99	0.99	0.99
10x18	LSTM	0.98	1.00	0.97	0.98
32x32x3	VGG16	0.99	0.99	0.99	0.99
32x732x3	Resnet50	0.99	1.00	0.98	0.99
75x75x3	Inception V3	0.99	1.00	0.99	0.99

The values reported in the aforementioned table were obtained during testing using sugarcane fields from Sargodha district and non-sugarcane pixels from Narowal fields. The test dataset consisted of approximately two thousand samples, with an equal number of samples from each class. The confusion matrix for each classification model, trained with a sample size of 10x18, is presented below.



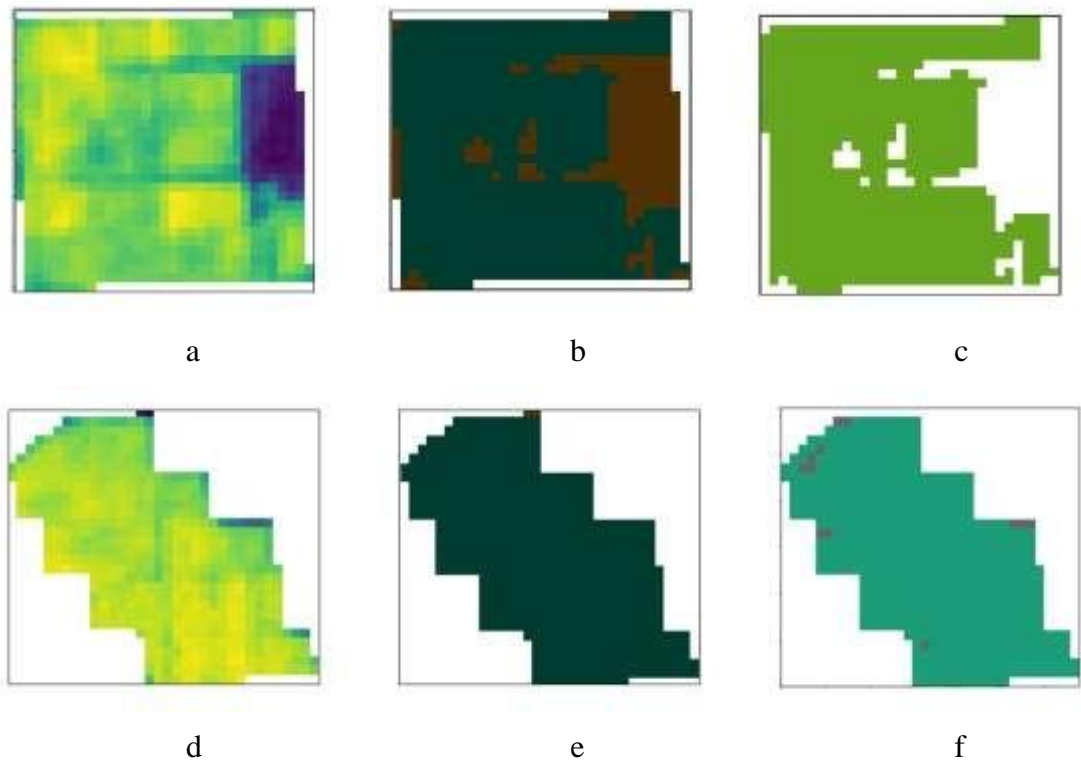
**Figure 5.2: Confusion matrix for the models with input sample size 10x18**

The table and confusion matrix provide a visual representation of the impressive classification results achieved by MLP, convnet, and LSTM using 2D time-series spectral feature maps for sugarcane classification. The accuracy of the models is evident from the classification maps, which visually depict the classified pixels. Each pixel in the classification map corresponds to the same location from the sentinel 2 cropped field imagery, highlighting the accuracy and consistency of the classification models.

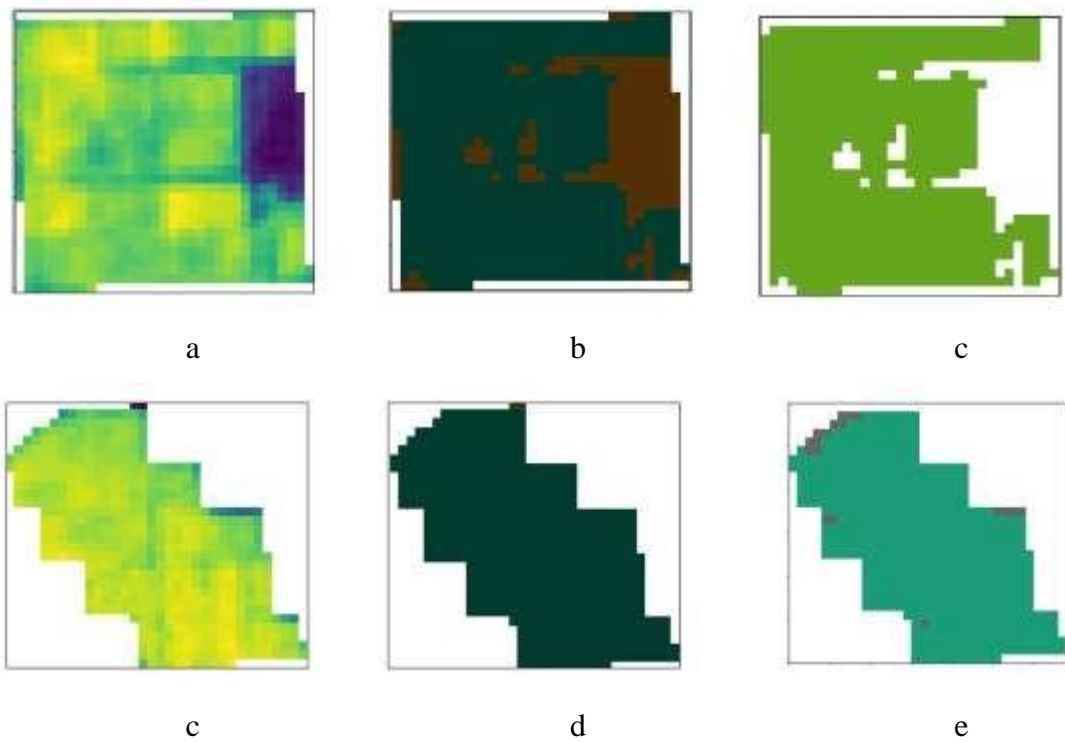


**Figure 5.3:** Classification model MLP; a) cropped sugarcane field b) abundance map c) classification map green pixel shows sugarcane grey shows non-sugarcane, d) non-sugarcane field, e) abundance map of green vegetation, f) classification map green shows non-sugarcane and grey shows sugarcane

From cropped field spectral imagery 2D timeseries spectral feature maps are created and classified as sugarcane or non-sugarcane pixel by testing on trained classification models. When arranged in the classification map it can be visualized whether the field is sugarcane or not.



**Figure 5.4:** Classification model Convnet; a) cropped sugarcane field b) abundance map c) classification map green pixel shows sugarcane grey shows non-sugarcane, d) non-sugarcane field, e) abundance map of green vegetation, f) classification map green shows non-sugarcane and grey shows sugarcane



**Figure 5.5:** Classification model LSTM; a) cropped sugarcane field b) abundance map c) classification map green pixel shows sugarcane grey shows non-sugarcane, d)

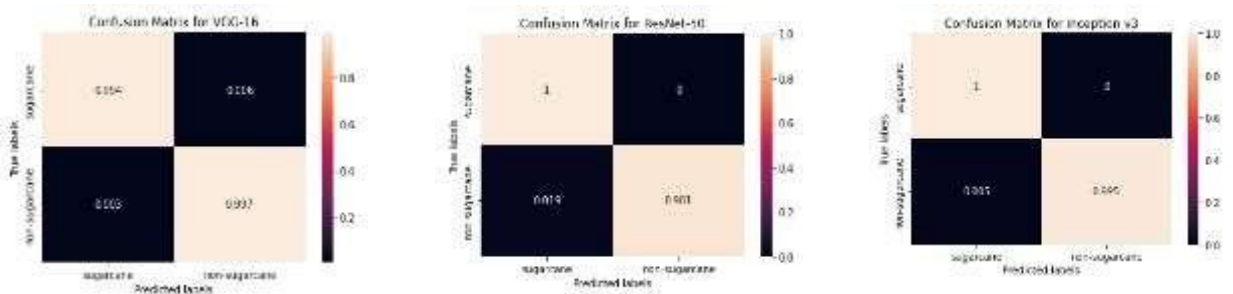
non-sugarcane field, e) abundance map of green vegetation, f) classification map  
 green shows non-sugarcane and grey shows sugarcane

In classifying sugarcane field all the three models accurately classified all pixels as the actual class, whereas in case of non-sugarcane some of the boundary pixels are classified incorrectly.

In order to explore more classification models, 2D Spectral maps are up sampled to cater the requirement of size which is minimum of 32x32x3. Single channel is stacked together thrice to make it three channel input.

The test results for all three CNN architectures show similar performance. Our classification accuracy for identifying sugarcane pixels among other crops such as wheat, rice, and corn is approximately 99.55% for VGG16, 99.05% for ResNet-50, and 99.75% for Inception v3. Notably, the results remain consistent regardless of whether transfer learning or parameter training is employed.

Figure 5.6 presents the confusion matrix for VGG16, ResNet-50, and Inception v3. However, ConvNext Tiny did not perform well in both training and testing phases, with an accuracy of approximately 51%. This suggests that ConvNext Tiny failed to effectively extract features from the 2D multispectral feature maps during training, resulting in unsatisfactory results during testing.



**Figure 5.6:** Confusion matrix for classification models VGG16, ResNet-50, Inception v3

### 5.1.3 Classification results for dataset combination 2

The results for the second dataset combination are not as impressive as those of the first combination. This is due to the high variance and noise present in this dataset combination, resulting in lower comparable accuracies compared to the first combination. Among the models, LSTM performed the best with an accuracy of 89.22%, indicating its ability to capture the time sequence dependencies in the data.

ResNet50 and VGG16 achieved similar performances of approximately 85% and 88%, respectively, as they were able to extract relevant features from different crops despite variations in location, sowing year, and crop stage during sentinel product selection. However, in the case of Inception V3, the input sample size should be 75x75x3. When the 10x18 image size is up-sampled using interpolation, a significant amount of noise is introduced into the data, causing the model to struggle in accurately fitting the test data. The lower accuracy compared to the first case indicates a high variance in the dataset. To reduce variance, additional data similar to the test data can be added to the training dataset, which helps in stabilizing the model parameters.

**Table 5.2: Training and test accuracy for dataset combination 2**

Input sample size	Architecture	Training Accuracy %	Test Accuracy %
32x32x3	VGG16	98.40	82.25
32x32x3	ResNet-50	99.01	82.24
75x75x3	Inception v3	98.11	64.23
32x32x3	LSTM	99.60	89.22
32x32x3	ConvNext(without data augmentation)	51.91	51.32
32x32x3	ConvNext (with data augmentation)	91.41	79.83

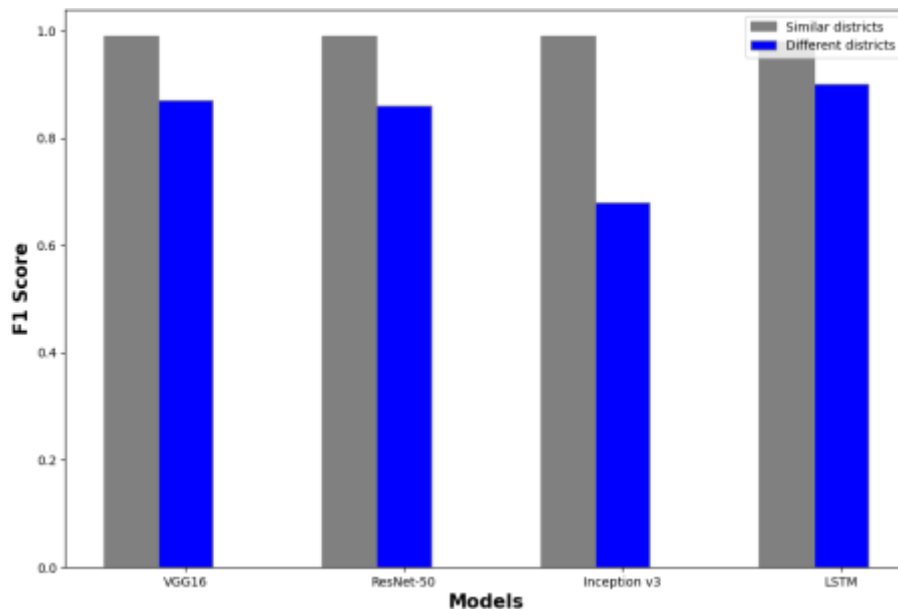
The confusion matrix of LSTM, VGG16, and ResNet-50 for this dataset combination is shown in figure 5.7. and evaluation metrics for all the classification models for dataset combination 2 are shown in table 5.3.

**Table 5.3: Evaluation metrics along with input sample size for dataset combination 2**

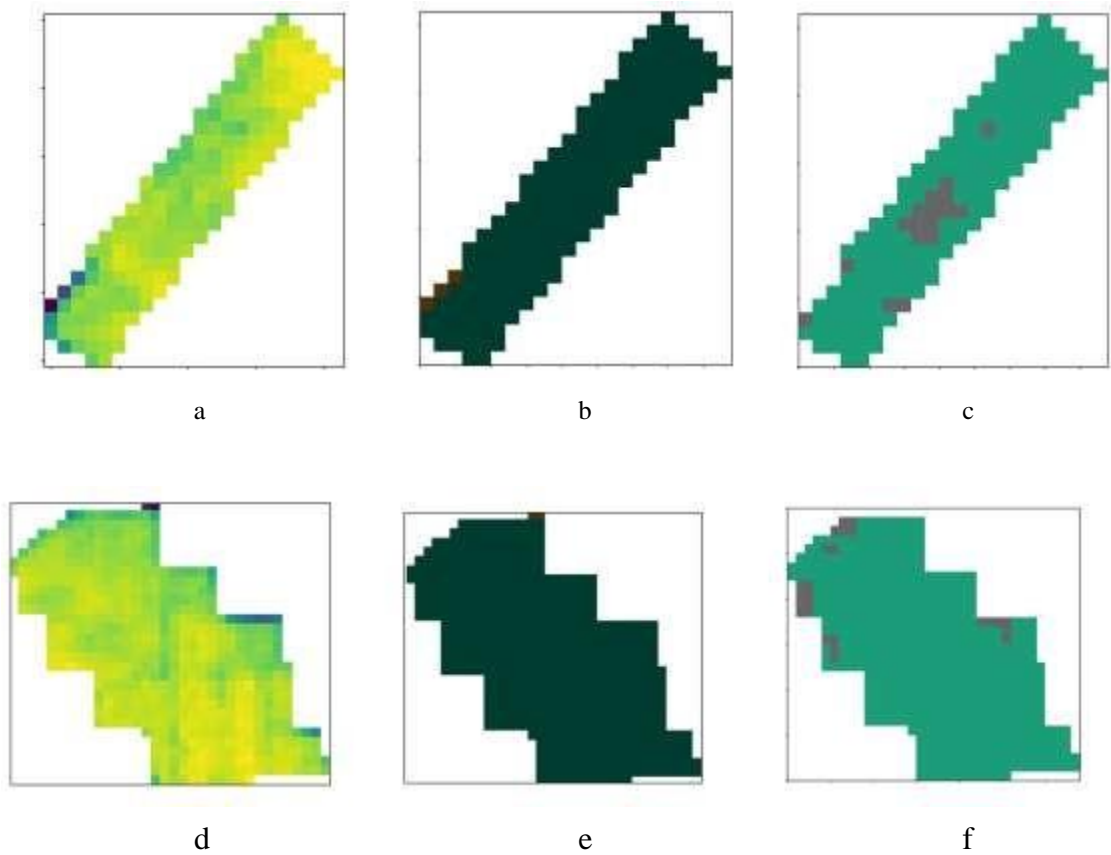
Input sample size	Architecture	Test Accuracy %	Precision	Recall	F1 Score
10x18	MLP	0.66	0.67	0.64	0.65
10x18	Simple	0.72	0.71	0.73	0.72

convnet					
10x18	LSTM	0.90	0.87	0.94	0.90
32x32x3	VGG16	0.88	0.85	0.91	0.87
32x32x3	ResNet-50	0.85	0.98	0.78	0.86
32x32x3	ConvNext	0.78	0.75	0.85	0.79
75x75x3	Inception v3	0.68	0.70	0.66	0.68

In both dataset combinations, ConvNext achieved an accuracy of approximately 51%, indicating its failure to effectively capture the features present in the training data. However, after applying data augmentation techniques, the training accuracy improved to 91% and the test accuracy increased to 79.83%. This suggests that the ConvNext model has the capability to perform well when the size of the dataset is increased. Figure 5.8 provides a qualitative comparison of F1 score for both the dataset combination.



**Fig. 5.8.** Performance of classification models on both dataset combinations i.e. (similar district and different district)



**Figure 5.9:** Classification model VGG16 for dataset combination 2; a) cropped sugarcane field b) abundance map c) classification map green pixel shows sugarcane grey shows non-sugarcane, d) non-sugarcane field, e) abundance map of green vegetation, f) classification map green shows non-sugarcane and grey shows sugarcane

Based on the results, it is evident that in order to enhance the robustness of our classifier, it is recommended to incorporate sugarcane and other crop datasets from various regions across Pakistan into our training data. This is important because crop characteristics can vary between different areas due to factors such as soil composition, crop varieties, and crop duration. Furthermore, increasing the number of products utilized will enable our model to learn a wider range of crop phenology attributes and gain a deeper understanding of crop characteristics overall.

## Chapter 6: Conclusion

Ensuring accurate information regarding the readiness of sugarcane fields for harvesting is essential to optimize yield and quality. Harvesting at the right maturity level maximizes sugar content, reduces losses due to overripening or delays, and provides valuable insights for long-term planning and market forecasting.

To address this, a deep learning-based framework is proposed for identifying ready-to-harvest sugarcane fields among other popular crops grown in Pakistan. The framework involves selecting Sentinel-2 Level 2A products based on NDVI time-series plots, followed by preprocessing and spectral feature extraction using SNAP. The temporal and multispectral features are organized into 2D images referred to as 2D time-series multispectral feature maps in this study. These feature maps are classified using multiple classification models.

The developed framework exhibits positive potential in distinguishing sugarcane feature maps from other major crop feature maps within the same district. Despite the challenges posed by dataset variability in location and dates, as well as the limited number of products used, our methodology excelled in feature extraction and pixel-based classification, yielding impressive results.

Experimental findings indicate that extracting essential features throughout the entire crop life cycle for each crop requires a quantity of ten time-series samples. Additional tests using five and seven time-series samples did not yield satisfactory results. Thus, utilizing ten time-series samples ensures sufficient multitemporal features.

The classification task primarily focuses on distinguishing sugarcane from crops like wheat, rice, and corn, which have shorter growth periods of 4-6 months compared to sugarcane's 10–12-month maturity cycle in Pakistan. To ensure fairness, an equal number of Sentinel-2 tiles were collected for all crops, resulting in a significant time gap for sugarcane. This methodology involves feature extraction and compilation of each pixel as an image, and it performs well even with MLP and simple Convnet models, achieving F1 scores of 0.94 and 0.99, respectively, when trained and tested within the same district. CNN architectures and LSTMs significantly contribute to recognizing sugarcane from other districts, where MLP



and simple Convnet models performed poorly. The sugarcane classification task achieves impressive results using LSTM, VGG16, and ResNet-50 models, with F1 scores of 0.90, 0.87, and 0.86, respectively.

To enhance the performance of the proposed framework, which is capable of identifying mature sugarcane fields across Pakistan, it is necessary to include data from most sugarcane-growing districts in the training dataset. Additionally, the choice of ten samples should cover all growing phases of sugarcane, including germination, tillering, grand growth, ripening, and maturation phases.

# BIBLIOGRAPHY

- [1] “Agriculture Statistics | Pakistan Bureau of Statistics.” <https://www.pbs.gov.pk/content/agriculture-statistics> (accessed Jun. 02, 2023).
- [2] S. Afghan *et al.*, “Economic Importance and Yield Potential of Sugarcane in Pakistan,” *Sugarcane - Its Products and Sustainability*, Feb. 2023, doi: 10.5772/INTECHOPEN.105517.
- [3] “THIS REPORT CONTAINS ASSESSMENTS OF COMMODITY AND TRADE ISSUES MADE BY USDA STAFF AND NOT NECESSARILY STATEMENTS OF OFFICIAL U.S. GOVERNMENT POLICY,” 2021.
- [4] M. Immitzer, F. Vuolo, and C. Atzberger, “First experience with Sentinel-2 data for crop and tree species classifications in central Europe,” *Remote Sens (Basel)*, vol. 8, no. 3, p. 166, 2016.
- [5] R. Sonobe, Y. Yamaya, H. Tani, X. Wang, N. Kobayashi, and K. Mochizuki, “Crop classification from Sentinel-2-derived vegetation indices using ensemble learning,” *J Appl Remote Sens*, vol. 12, no. 2, p. 26019, 2018.
- [6] M. G. Maponya, A. Van Niekerk, and Z. E. Mashimbye, “Pre-harvest classification of crop types using a Sentinel-2 time-series and machine learning,” *Comput Electron Agric*, vol. 169, p. 105164, 2020.
- [7] A. Chakhar, D. Hernández-López, R. Ballesteros, and M. A. Moreno, “Improving the accuracy of multiple algorithms for crop classification by integrating sentinel-1 observations with sentinel-2 data,” *Remote Sens (Basel)*, vol. 13, no. 2, p. 243, 2021.
- [8] L. Piedelobo *et al.*, “Scalable pixel-based crop classification combining Sentinel-2 and Landsat-8 data time series: Case study of the Duero riverbasin,” *Agric Syst*, vol. 171, pp. 36–50, 2019.
- [9] A. Chakhar, D. Ortega-Terol, D. Hernández-López, R. Ballesteros, J. F. Ortega, and M. A. Moreno, “Assessing the accuracy of multiple classification algorithms for crop classification using Landsat-8 and Sentinel-2 data,” *Remote Sens (Basel)*, vol. 12, no. 11, p. 1735, 2020.

- [10] M. Belgiu and O. Csillik, "Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis," *Remote Sens Environ*, vol. 204, pp. 509–523, 2018.
- [11] S. Moharana, B. Kambhammettu, S. Chintala, A. S. Rani, and R. Avtar, "Spatial distribution of inter-and intra-crop variability using time-weighted dynamic time warping analysis from Sentinel-1 datasets," *Remote Sens Appl*, vol. 24, p. 100630, 2021.
- [12] Y. Zheng *et al.*, "Development of a Phenology-Based Method for Identifying Sugarcane Plantation Areas in China Using High-Resolution Satellite Datasets," *Remote Sens (Basel)*, vol. 14, no. 5, p. 1274, 2022.
- [13] V. Mazzia, A. Khaliq, and M. Chiaberge, "Improvement in land cover and crop classification based on temporal features learning from Sentinel-2 data using recurrent-convolutional neural network (R-CNN)," *Applied Sciences*, vol. 10, no. 1, p. 238, 2019.
- [14] Q. Li, J. Tian, and Q. Tian, "Deep Learning Application for Crop Classification via Multi-Temporal Remote Sensing Images," *Agriculture*, vol. 13, no. 4, p. 906, 2023.
- [15] G. Siesto, M. Fernández-Sellers, and A. Lozano-Tello, "Crop Classification of Satellite Imagery Using Synthetic Multitemporal and Multispectral Images in Convolutional Neural Networks," *Remote Sens (Basel)*, vol. 13, no. 17, p. 3378, 2021.
- [16] U. Rauf *et al.*, "A new method for pixel classification for rice variety identification using spectral and time series data from Sentinel-2 satellite imagery," *Comput Electron Agric*, vol. 193, p. 106731, 2022.
- [17] S. Ofori-Ampofo, C. Pelletier, and S. Lang, "Crop type mapping from optical and radar time series using attention-based deep learning," *Remote Sens (Basel)*, vol. 13, no. 22, p. 4668, 2021.
- [18] Y. Wang, Z. Zhang, L. Feng, Y. Ma, and Q. Du, "A new attention-based CNN approach for crop mapping using time series Sentinel-2 images," *Comput Electron Agric*, vol. 184, p. 106090, 2021.
- [19] "Open Access Hub." <https://scihub.copernicus.eu/> (accessed Jun. 02, 2023).
- [20] "NCRA | National Centre of Robotics and Automation." <https://ncra.org.pk/> (accessed Jun. 02, 2023).

- [21] C. J. Tucker, “Red and photographic infrared linear combinations for monitoring vegetation,” *Remote Sens Environ*, vol. 8, no. 2, pp. 127–150, 1979.
- [22] B.-C. Gao, “Normalized difference water index for remote sensing of vegetation liquid water from space,” in *Imaging Spectrometry*, 1995, pp. 225–236.
- [23] A. J. Richardson and C. L. Wiegand, “Distinguishing vegetation from soil background information,” *Photogramm Eng Remote Sensing*, vol. 43, no. 12, pp. 1541–1552, 1977.
- [24] A. R. Huete, “A soil-adjusted vegetation index (SAVI),” *Remote Sens Environ*, vol. 25, no. 3, pp. 295–309, 1988.
- [25] R. F. Williams, “The physiology of plant growth with special reference to the concept of net assimilation rate,” *Ann Bot*, vol. 10, no. 37, pp. 41–72, 1946.
- [26] S. Sánchez, G. Martín, A. Plaza, and C.-I. Chang, “GPU Implementation of Fully Constrained Linear Spectral Unmixing for Remotely Sensed Hyperspectral Data Exploitation,” 2010, doi: 10.1117/12.860775.
- [27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, 2009, pp. 248–255.
- [28] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [29] “Everything you need to know about VGG16 | by Great Learning | Medium.” <https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918> (accessed Jun. 08, 2023).
- [30] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [31] “The Annotated ResNet-50. Explaining how ResNet-50 works and why... | by Suvaditya Mukherjee | Towards Data Science.” <https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758> (accessed Jun. 08, 2023).

- [32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.
- [33] “Architecture of the Inception V3 module [41]. | Download Scientific Diagram.” [https://www.researchgate.net/figure/Architecture-of-the-Inception-V3-module-41\\_fig19\\_331980242](https://www.researchgate.net/figure/Architecture-of-the-Inception-V3-module-41_fig19_331980242) (accessed Jun. 08, 2023).
- [34] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [35] “LSTM Recurrent Neural Networks — How to Teach a Network to Remember the Past | by Saul Dobilas | Towards Data Science.” <https://towardsdatascience.com/lstm-recurrent-neural-networks-how-to-teach-a-network-to-remember-the-past-55e54c2ff22e> (accessed Jun. 08, 2023).
- [36] A. Vaswani *et al.*, “Attention is all you need,” *Adv Neural Inf Process Syst*, vol. 30, 2017.
- [37] A. Dosovitskiy *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.
- [38] Z. Liu *et al.*, “Swin transformer: Hierarchical vision transformer using shifted windows,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10012–10022.
- [39] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, “A convnet for the 2020s,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11976–11986.
- [40] “Architecture of the ConvNeXt network, a four-stage feature hierarchy,... | Download Scientific Diagram.” [https://www.researchgate.net/figure/Architecture-of-the-ConvNeXt-network-a-four-stage-feature-hierarchy-was-built-to\\_fig2\\_365870304](https://www.researchgate.net/figure/Architecture-of-the-ConvNeXt-network-a-four-stage-feature-hierarchy-was-built-to_fig2_365870304) (accessed Jun. 08, 2023).
- [41] “Google Earth Pro.” Google LLC, 2022.
- [42] S. S. and C.-S. Brockmann Consult, “SNAP.” 2014.
- [43] “What is a Confusion Matrix in Machine Learning?” <https://www.simplilearn.com/tutorials/machine-learning-tutorial/confusion-matrix-machine-learning> (accessed Jun. 08, 2023).