

Crop type classification using multi-sensor satellite image time series data and attention-based deep learning technique: a case study of southern India

Shivaprakash Yaragal

2025
Department of
Physical Geography and Ecosystem Science
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden



Shivaprakash Yaragal (2025).

Crop type classification using multi-sensor satellite image time series data and attention-based deep learning technique: a case study of southern India

Grödklassificering med hjälp av multisensor satellitbildstidsseriedata och uppmärksamhetsbaserad djupinlärningsteknik: En fallstudie från södra Indien

Master degree thesis, 30 credits in *GIS and Remote Sensing*
Department of Physical Geography and Ecosystem Science, Lund University

Level: Master of Science (MSc)

Course duration: *January 2025 until June 2025*

Disclaimer

This document describes work undertaken as part of a program of study at the University of Lund. All views and opinions expressed herein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

Crop type classification using multi-sensor satellite image time series data and attention-based deep learning technique: a case study of southern India

Shivaprakash Yaragal

Master thesis, 30 credits, in *GIS and Remote Sensing*

Prof. Ali Mansourian

GIS Center, Department of Physical Geography and Ecosystem Science
Lund University

Dr Rachid Oucheikh

GIS Center, Department of Physical Geography and Ecosystem Science
Lund University

Exam committee:

Dr Zheng Duan

Department of Physical Geography and Ecosystem Science , Lund University

Wan Ni Lin

Department of Physical Geography and Ecosystem Science, Lund University

Acknowledgement

I am deeply grateful to the Lund University Global Scholarship (LUGS) program for awarding me a full scholarship to pursue my master's degree at Lund University. I also thank the Crafoord Foundation for their travel grant, which enabled essential fieldwork in India that forms the foundation of this thesis.

My sincere thanks go to my supervisors, Prof. Ali Mansourian and Dr. Rachid Oucheikh, whose guidance, mentorship, and especially Prof. Ali's belief in the project were pivotal. Their constructive feedback and unwavering support - particularly the Zoom meetings held by Dr. Rachid - have greatly shaped this work. I also wish to express my gratitude to Dr. Zheng Duan, who, in his role as examiner, provided valuable insights that significantly enhanced the quality and clarity of this thesis.

This work would not have been possible without the support of my family in India. My wife, Shilpashree Nadagaudar, provided unwavering support throughout every phase of this journey, and my father, Siddanna S Yaragall, stood by me with constant encouragement. The support of my entire family during long field visits and travel was a continuous source of strength and comfort.

Special thanks to Dr. Shashikant Evoor, whose expertise in crop knowledge and generous guidance during data collection played a critical role. I am also grateful to the families of Basavaraj Jagatkal and Malanagauda Patil who warmly hosted me during my fieldwork. My heartfelt appreciation extends to all the villagers who provided directions and assistance, making the data collection process possible.

I am also sincerely thankful to my opponent, Mengjie Yu, for his valuable feedback and thoughtful suggestions on my thesis writing. To my close friends Elias Haglund and Mingyuan Zhang, who supported me in various ways - thank you for being there during critical moments. I also acknowledge the contributions of Stella Ofori-Ampofo, whose work and suggestions were valuable to my project.

In this work, I benefited greatly from Windsurf Editor, an AI-powered coding and chat environment that helped solve several programming challenges efficiently. I also acknowledge Esri's platform, which provided crucial tools for collecting and managing spatial data during the field work.

To all who contributed to this thesis — in ways big or small — I offer my deepest gratitude.

I would like to close with the words of Basavanna, a 12th-century philosopher and social reformer, whose teachings continue to inspire me:

ಕಾಯಕವೇ ಕೈಲಾಸ — Work is Worship

Abstract

Crop type classification is essential for effective agricultural monitoring in India. Performing crop type classification in the diverse Indian agriculture setting is a challenging task. This thesis leverages transformer-based deep learning model, specifically the Pixel-Set Encoder Temporal Attention Encoder (PSETAE), to classify ten crop types of rabi season (2024 - 25) using Satellite Image Time Series (SITS) data. The thesis provides an approach that integrates mobile-based field surveys, Google Earth Engine (GEE) based data extraction and preprocessing, and model implementation using PyTorch. Field data involving polygons with the crop label were collected from two regions of Karnataka, India, using ArcGIS Field Maps. The polygons served as the basis for extracting SITS data, which were then utilized for crop classification. Multiple SITS data - Sentinel-1, Sentinel-2, and Harmonized Landsat Sentinel-2 datasets (HLSS30 and HLSL30) - and data fusion methods were evaluated using PSETAE crop type classification model. The PSETAE model achieved high classification accuracies for Sentinel-2, followed by HLSS30 and HLSL30 and Sentinel-1 achieved least. Fusion Methods achieved accuracy between 94-95 % which is slightly less than Sentinel-2. Spatial generalization tests revealed a remarkable inversion of results suggesting that Synthetic Aperture Radar (SAR) data from Sentinel-1 captures more transferable information about crops than Sentinel-2. Both HLSS30 and HLSL30 performed well in the test area but failed to maintain their performance during generalization. While overall generalization accuracy was low, strong F1 scores for chickpea (0.92), wheat (0.80), and sugarcane (0.79) from Sentinel-1, and fruit plantations from HLSL30 (0.82) and Statistical Average Fusion (0.83), highlight the potential of Sentinel-1, data fusion methods and model ensembling to enhance operational crop type classification.

Keywords: agriculture; crop type classification; transformer-based deep learning; PSETAE; SITS; Sentinel-1; Sentinel-2; HLS products; fusion methods; ArcGIS ; GEE; PyTorch

Table of contents

1.	Introduction.....	1
1.1	Agriculture monitoring in India	2
1.2	Crop type mapping.....	2
1.3	Research objectives.....	4
1.4	Study scope	4
1.5	Thesis outline	4
2.	Literature Review.....	5
2.1	Satellite data.....	5
2.1.1	Sentinel-1	6
2.1.2	Sentinel-2	6
2.2	Satellite data augmentation and data fusion.....	7
2.3	Machine learning and deep learning in crop type mapping.....	8
2.3.1	Transformer based model	9
3.	Method and Materials	11
3.1	Study area.....	11
3.2	Methodology	12
3.2.1	Mobile based data collection	13
3.2.2	Application design	13
3.2.3	Survey planning	14
3.2.4	Field data characteristics.....	15
3.3	Data processing using GEE	16
3.4	PSETAE model.....	18
3.4.1	Pixel-Set Encoder.....	18
3.4.2	Temporal Attention Encoder.....	19
3.5	Implementation	19
3.5.1	Data extraction	19
3.5.2	Hyperparameter tuning	19
3.5.3	Experimental design.....	20
3.5.4	Choice of fusion methods	20
3.5.5	Evaluation metrics	22
4.	Results.....	24
4.1	Evaluation of model learning.....	24
4.2	Comparative analysis using evaluation metrics.....	25

4.3	Performance during initial growing period.....	30
4.4	Model inferencing on Region B.....	31
5.	Discussion.....	35
5.1	Interpretation of the results	35
5.1.1	Class specific interpretation.....	36
5.2	Implications and limitations.....	37
6.	Conclusion and Future Research	40
	References.....	41
	Appendix.....	A
	Appendix A: Survey planning and field survey.....	A
	Appendix B: Observed crop mixing	D
	Appendix C: Cloud cover statistic and filtering	F
	Appendix D: PSETAE equations.....	H

List of Abbreviations

AER	Agro-eco Regions
AESR	Agro-eco Sub-Regions
AGOL	ArcGIS Online
ARD	Analysis Ready Data
CNN	Convolutional Neural Networks
EVI	Enhanced Vegetation Index
ESA	European Space Agency
EOS	Earth Observing System
FASAL	Forecasting Agricultural Output using Space, Agro-meteorology, and Land-based Observations
GEE	Google Earth Engine
GRD	Ground Range Detected
GRU	Gated Recurrent Units
HLS	Harmonized Landsat and Sentinel-2
IAS	Improvement of Agricultural Statistics
ISRO	Indian Space Research Organization
JPSS	Joint Polar Satellite System
LISS	Linear Imaging Self Scanner
LSTM	Long Short-Term Memory
LSWI	Land Surface Water Index
MODIS	Moderate Resolution Imaging Spectroradiometer
MSI	Multispectral Imager
MSS	Multispectral Scanner
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
OLI	Operational Land Imager
PSETAE	Pixel-set Encoder Temporal Attention Encoder
RF	Random Forest
RNN	Recurrent Neural Networks
SAR	Synthetic Aperture Radar
SITS	Satellite Image Time Series
SLC	Single Look Complex
SVM	Support Vector Machines
TM	Thematic Mapper
VH	Vertical transmit and Horizontal receive
VNIR	Visible and Near Infrared Region
VV	Vertical transmit and Vertical receive
XGB	eXtreme Gradient Boosting

1. Introduction

Sustainable Development, a defining concept of our time, is both a framework for understanding global challenges and a collective strategy for addressing them. The current world is facing escalating global challenges, including population growth, rising global temperature, economic distress, and resource constraints. Ensuring adequate availability of food, part of SDGs 2: Zero Hunger, is a critical challenge of food security amidst these escalating challenges (Brown, 2014). Various reports of 2024 highlighted that between 713 to 757 million people may have faced hunger in 2023 (FAO, IFAD, UNICEF, WFP, & WHO, 2024). The world population is projected to reach up to 10.83 billion by 2082 (World Population Prospects, United Nations, 2024), this will add to the existing pressure on the resource there by aggravating the situation of food availability. To meet the demand of increasing population, there is increasing demand for food security monitoring systems involving monitoring of crop type, crop health, cropping date, vegetation, yield estimation, damage assessment and many more (Brown, 2014).

Agricultural monitoring has undergone significant change over the years. Traditional methods of monitoring involved manual methods which were labor intensive, time consuming and resource dependent (UNOOSA & FAO, 2025; Victor et al., 2025). Over the years, with the rapid development of technology, there has been a gradual move towards building a monitoring system that integrates real time mobile base data collection, satellite imagery, cloud technology and machine learning (UNOOSA & FAO, 2025). Famine Early Warning Systems Network (FEWS NET), a United States initiative, pioneered the use of satellite remote sensing to establish a system to monitor agriculture and related food security (Brown, 2014). FEWS NET, as it stands today, is a full-fledged decision support system (DSS) (Senay et al., 2020). Similarly, Group on Earth Observations Global Agricultural Monitoring (GEOGLAM), an international program was established to leverage satellite data in improving crop monitoring, yield forecasting, early drought warning and better agricultural decision-making (UNOOSA & FAO, 2025; Becker-Reshef et al., 2023).

Satellite data plays a central role in agricultural monitoring for a very large area. National Aeronautics and Space Administration's (NASA) programs like Landsat, Earth Observing System (EOS), Joint Polar Satellite System (JPSS), etc. have contributed immensely to agricultural monitoring at regional and global (Song et al., 2021; Brown, 2014). Similarly, Sentinel satellites under the Copernicus program managed by European Space Agency (ESA) have paved the way for better, efficient and high frequency monitoring (Song et al., 2021; Copernicus. (n.d.)). The Moderate Resolution Imaging Spectroradiometer (MODIS) sensor, with spatial resolution ranging from 250 to 1000 meters onboard Aqua and Terra (EOS program), the Landsat satellite series with 30-meter resolution imaging sensors, and the Sentinel satellite series, especially Sentinel-1 (10-meter resolution) and Sentinel-2 (10-, 20-, and 60-meter resolutions), are popular for crop monitoring (Song et al., 2021; Sheykhmousa et al., 2020; Alami Machichi et al., 2023; Victor et al., 2025). These satellites are widely used because of their high revisit frequency (weekly or

biweekly), moderate-to-coarse spatial resolution (10 to 250 meters), global coverage, and free accessibility (Song et al., 2021). While high spatial resolution imagery (below 10 meters) can be used for agricultural monitoring (Rao et al., 2021; Wang et al., 2020), it is economically and computationally impractical for continental-scale (like region of India) monitoring.

1.1 Agriculture monitoring in India

Agriculture, being a complex system in India, plays a vital role. Cropland in India covers 154 million hectares, accounting for 42.8% of India's territorial area and contributes 18.4% to India's Gross Value Added (GVA) (Government of India. 2023, Agriculture Statistics at a glance 2023-24). Cropland is spread across 20 agro-eco regions (AER), which are further divided into 60 agro-eco subregions (AESR), making it one of the most complex agriculture landscapes (Gajbhiye & Mandal, 2000). With more than 100 crops, including 25 major ones (Sivaraj et al., 2013) and three cropping seasons - kharif, rabi and zaid - the complexity of Indian agriculture becomes even more noticeable.

Monitoring complex agricultural systems requires multifaceted efforts from government organizations. The Improvement of Agricultural Statistics (IAS) scheme and the Forecasting Agricultural Output using Space, Agro-meteorology, and Land-based Observations (FASAL) program are key initiatives in this direction. The IAS scheme conducts crop-cutting experiments in over 10,000 villages to generate reliable seasonal estimates of crop area and production. Similarly, the FASAL program integrates remote sensing, ground observations, and meteorological data, using machine learning models developed by Indian Space Research Organization (ISRO) to forecast crop area and yield at district, state, and national levels for major crops. Together, these programs support timely and accurate agricultural monitoring across India.

1.2 Crop type mapping

Crop type mapping is the process of identifying and classifying different types of crops in the agricultural field using remote sensing techniques (Alami Machichi et al., 2023). It provides the spatial distribution of types of crops in study area (Becker-Reshef et al., 2023), making it a critical step in agriculture monitoring. Agriculture monitoring includes early warning of crop yield prediction (Alami Machichi et al., 2023; Victor et al., 2025, Singh et al., 2022), damage assessment (Becker-Reshef et al., 2023; Victor et al., 2025), production forecast (Koppaka & Moh, 2020; Prashant & Michael, 2020), agriculture insurance management (Prashant & Michael, 2020) and agriculture planning and support system (Prashant & Michael, 2020; Singh et al., 2022).

Crop mapping is a crucial tool for supporting regional agricultural planning and policymaking. Crop mapping helps in understanding shifting cropping patterns, managing over-production and under-production, and maintaining price stability, thereby contributing to regional food security (Brown, 2014). A recent review by You and Sun (2022) highlights how crop mapping provides policymakers with evidence-based information for addressing targeted rural development, resource allocation, and thereby reducing regional disparities. Similarly, Mahlayeye et al. (2022)

discuss how crop mapping aids in analyzing regional cropping patterns, which in turn supports the development of incentive-based agricultural programs, tracks policy effectiveness, and contributes to regional development.

Implementation of effective crop type mapping in India faces several challenges. The primary challenge is the predominance of small landholdings (Prashant & Michael, 2020; Rao et al., 2021; Singh et al., 2022; Ayushi & Buttar, 2024). According to Government of India's Agriculture Statistics at a Glance 2023-24, nearly 86% of total landholdings are under two hectares. Such small landholding makes it difficult to obtain enough pixels for accurate crop classification using medium-resolution satellite imagery (Karale et al., 2014; Ayushi & Buttar, 2024), leading to higher misclassifications (Singh et al., 2013; Rao et al., 2021). Furthermore, a single landholding may consist of multiple fields (Graesser & Ramankutty, 2017), and individual fields often contain more than one crop. This practice, known as multi-cropping, is widely adopted in many parts of South Asia (Wang et al., 2020). This heterogeneity presents a second major challenge especially during data collection, as field workers often struggle to account for it, leading to data mislabeling that compromises accuracy of classification.

The third major challenge is the non-availability of open data for doing research and development. Though government initiatives like FASAL and KrishiMapper App have contributed to the collection of polygon-based ground truth data, there has been no effort to open this data for research. The final major challenge is the lack of publicly available free high-resolution satellite imagery to accurately map small agricultural plots (Rao et al., 2021). Although Linear Imaging Self Scanner-IV (LISS-IV) data from the Government of India offers a 5.8-meter resolution (ISRO, 2016), it is not freely available and lacks the temporal repeatability required for effective crop type mapping. Additionally, LISS-IV is limited by three spectral bands in the Visible and Near Infrared Region (VNIR) (ISRO, 2016), which reduces its ability to differentiate between various crop types effectively, thereby limiting its suitability for crop mapping.

Given the challenges discussed earlier, such as the non-availability of open ground truth data, restricted access to high-resolution satellite imagery, and small landholdings, there is a need to explore new methods for classifying crop types using satellite data that leverage dense, openly available multi-sensor Satellite Image Time Series (SITS), and use limited ground truth to achieve good crop type classification. Recent advancements in deep learning, particularly the attention-based approach used in Pixel-Set Encoder Temporal Attention Encoder (PSETAE), offer the opportunity to make use of SITS data for crop classification. It also provides methods to explore data integration of SITS such as Sentinel-1 and Sentinel-2 to further enhance accuracy. In addition, the newly developed Harmonized Landsat and Sentinel-2 (HLS) dataset remains untested in the Indian context for crop type classification, providing an opportunity for novel investigation using PSETAE. Finally, assessing the spatial transferability of the model represents a critical step toward building scalable and generalizable crop mapping solutions. These above considerations form the basis for this master's thesis.

1.3 Research objectives

The aim is to explore transformer based deep-learning approach for crop type classification using multi-sensor satellite image time series (SITS) in diverse agricultural landscapes of India.

This master's thesis aims to answer the following research questions:

- 1) Does integrating Sentinel-1 and Sentinel-2 through attention-based model (PSETAE) significantly improve crop classification accuracy compared to using either sensor alone, and which fusion methods perform better for India's diverse agricultural landscapes?
- 2) To what extent does combining Sentinel-1 and Sentinel-2 data enhance classification accuracy compared to using Harmonized Landsat and Sentinel-2 (HLS) dataset?
- 3) How well does the PSETAE model demonstrate spatial transferability when trained on one rural region and tested in a nearby rural region?

1.4 Study scope

The thesis uses PSETAE model to carry out crop type classification using mobile survey ground truth and satellite time series data. Study is confined to the rabi cropping season and considers only major crop types for developing the crop type classification model. The satellite datasets used in this study are limited to Sentinel-1, Sentinel-2, and the Harmonized Landsat and Sentinel-2 (HLS) products; no other satellite data sources were utilized. The study does not include comparisons between PSETAE and other machine learning or deep learning models. Data collection was restricted to two specific regions of Karnataka, India, which are described in detail in the study area section.

1.5 Thesis outline

This thesis consists of 6 chapters. After this introductory chapter that describes technological development in India and research objectives, chapter 2 review the literature, highlighting freely available satellite datasets, machine learning and deep learning techniques used in crop type classification. Selection of study areas, data collection and methodology are described in chapter 3. Chapter 3 also presents model implementation, experimental design, and choice of fusion methods. Results are depicted in chapter 4. Chapter 5 discusses the results and chapter 6 concludes and suggests future research directions on crop type classification.

2. Literature Review

The literature review details the availability of various satellite datasets for crop classification, followed by data augmentation and the fusion of multi-sensor data. The section then discusses machine learning and deep learning techniques applied to crop type classification, concluding with transformer-based models for crop type classification.

2.1 Satellite data

The usage of satellite data in crop type classification has evolved over the years. The first attempt at crop type classification in India was carried out in the Karnal district of Haryana State using Landsat Multispectral Scanner (MSS) data in 1985 (Singh et al., 2013). On similar line, an effort to develop a methodology for mapping crops such as wheat, rice, sorghum, cotton, and rapeseed, was made in the 1990s (Singh et al., 2013). Dadhwal et al. (2002), in their review paper, highlight the use of Landsat MSS, Landsat Thematic Mapper (TM), and Indian Remote Sensing (IRS)'s LISS-I and LISS-II for crop type classification, achieving an accuracy between 87% and 90%. Similarly, Kawarkhe and Musande (2014) used LISS-III derived indices to map cotton with the accuracy of around 85%. Singh et al. (2013) compared the accuracy of LISS-III and LISS-IV and concluded that LISS-IV, with a 5-meter resolution, provided better results than LISS-III, which has a 23.5-meter resolution.

MODIS and its products have been used for mapping certain crops at regional and continental levels. Teluguntla et al. (2015) utilized MODIS-derived indices such as the Normalized Difference Vegetation Index (NDVI), Land Surface Water Index (LSWI), and Enhanced Vegetation Index (EVI) to map rice fields in the Krishna river basin, achieving an accuracy of 78%. Similarly, Gumma et al. (2016) used MODIS's 250-meter, 16-day NDVI composite to map rice-fallow lands during the rabi season in a sub-continental study covering India, Pakistan, Nepal, Bangladesh, and Sri Lanka, with an achieved accuracy of 82%. Both Teluguntla et al. (2015) and Gumma et al. (2016) point out that MODIS's lower resolution is a key factor contributing to lower accuracy and recommend using higher-resolution data, such as Landsat.

The Landsat satellite series is by far the longest-serving satellite program, operating continuously from 1972 to the present, beginning with Landsat-1 and extending to the most recent mission, Landsat-9 (US Geological Survey, 2015). With its extensive temporal coverage and 30-meter spatial resolution, Landsat has been one of the most widely used satellite datasets for classification activities (Davidson et al., 2017; Asam et al., 2022; Sheykhmousa et al., 2020; Alami Machichi et al., 2023). The Landsat satellite series has facilitated the development of medium-resolution (30-meters) crop type mapping in regions characterized by large field sizes, such as the United States of America (USA) (USDA NASS, 2024) and Canada (Fisette et al., 2013). In India, Karale et al. (2014) and Paul and Kumar (2019) used Landsat-8 data for crop classification, achieving accuracies of 88.49% and 80.96%, respectively. Paul and Kumar (2019) further concluded that

Sentinel-2 with 10-meter resolution provides better crop classification accuracy compared to the Landsat-8 dataset.

Monitoring agriculture during the monsoon and post-monsoon season in India is highly challenging. During a good rainy season (referred to as the active monsoon season) most parts of India experience more than 70% cloud cover (Rana & Sathiyamoorthy, 2018). Monitoring during this period, which typically lasts from June to November, using optical remote sensing is particularly difficult (Defourny et al., 2019; Asam et al., 2022; Vijayan et al., 2023; Alami Machichi et al., 2023). However, Vijayan et al. (2023) demonstrated that it is possible to map crops even during the cloudy season using SAR data, achieving 94–95% accuracy in classifying rice fields in the Palakkad district.

2.1.1 Sentinel-1

Sentinel-1, a satellite mission by the ESA, collects data in C-band SAR with a revisit time of 12 days (Copernicus SentiWiki, n.d.). Currently, the mission has Sentinel-1A (launched on 3 April 2014) and Sentinel-1C (launched on 5 December 2024), with Sentinel-1B (launched on 25 April 2016) out of service from August 2022. Sentinel-1 collection comprising of Single Look Complex (SLC) and ground range detected (GRD) products can be used for crop monitoring. Typically, if the SAR intensity is preferred, then the GRD is the only product needed. In GRD products, cross-polarized VH (Vertical transmit and Horizontal receive) backscatter is sensitive to vegetation density and structure, while the VH/VV (Vertical transmit and vertical receive) polarization ratio - leveraging both VH and VV intensities - provides enhanced sensitivity to vegetation phenology and moisture content (Copernicus SentiWiki, n.d.).

Mandal et al. (2018) developed a processing chain called Sen4Rice for differentiating early and late transplanted rice using Sentinel-1 GRD dataset. Similarly, Nagendram et al. (2023) used Sentinel-1 data to map paddy cropland in Guntur district, Andhra Pradesh, India. A review by Song et al. (2021) highlights that Sentinel-1 data significantly enhances crop classification by enabling the retrieval of crop biophysical variables, thereby improving overall classification accuracy. The study also emphasizes that Sentinel-1 increases the number of observations during cloudy periods, making it a valuable tool for agricultural monitoring in monsoon-affected regions. Given these advantages, Sentinel-1 data presents an entirely new dataset with untapped potential for experimentation in crop type mapping, particularly in cloud-prone regions like southern parts of India.

2.1.2 Sentinel-2

Sentinel-2 is a multi-spectral, medium-resolution wide swath optical sensor having 13 spectral bands with four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution (Copernicus SentiWiki, n.d.). Three satellites of this mission, i.e. Sentinel-2A (launched on 23 June 2015), Sentinel-2B (launched on 7 March 2017) and Sentinel-2C (launched on 5 September 2024), are operational now, with Sentinel-2A on extension campaign. With two satellites in the orbit the revisit is every five days (Copernicus SentiWiki, n.d.). Because of its high revisit, good

resolution, and analysis ready products, Sentinel-2 is one of the sorts after satellite data for land monitoring and crop classification (Alami Machichi et al., 2023). Sentinel-2 as a standalone data has been used in crop classification as various bands helps to differentiate between the crops achieving good accuracy (Singh et al., 2022). It is also used along with other high resolutions such as Planet satellite data (Rao et al., 2021), making Sentinel-2 one of the preferred candidates for experimenting with combined dataset.

2.2 Satellite data augmentation and data fusion

Capturing the growth stages is crucial for crop classification. Skilled farmers or agricultural experts can monitor these stages through regular field visits. However, in computer vision, making machines understand phenological changes is challenging, requiring time-series data of crop growth. This challenge increases when dealing with satellite imagery. While sub-meter resolution satellites may capture changes in color and texture from above, detecting such variations using medium-resolution (10 to 30 meters) satellite data is significantly more difficult (Geirhos et al., 2019). To overcome these limitations, researchers have employed feature engineering and data augmentation techniques. Asam et al. (2022) demonstrated that NDVI is more effective in identifying crop types compared to the blue, green, and red bands of the Sentinel-2 dataset. Additionally, simple ratios, vegetation indices, water indices, and moisture indices have been incorporated into existing satellite bands to improve crop classification accuracy (Karale et al., 2014; Koppaka & Moh, 2020; Rao et al., 2021; Yan & Ryu, 2019).

Augmenting data temporally with time-series imagery has shown remarkable improvements over single date imagery (Alami Machichi et al., 2023; Blickensdörfer et al., 2022). Gumma et al. (2016) and Mandal et al. (2018) used MODIS and Sentinel-1 time series to work on rice. Ayushi & Buttar (2024) used NDVI along with temporal interpolation to capture crop growth stages, demonstrating the effectiveness of temporal augmentation in crop classification.

Capturing spectral signatures, growth stages, and temporal information are crucial for distinguishing crops using satellite data. Imagery provides spectral characteristics, SAR captures structural information, and time-series data reveal how both evolve over time. This multi-source augmentation enhances the available information, thereby improving the accuracy of crop type classification. Hamidi et al. (2024) demonstrated that a combined time-series dataset of Sentinel-2 and Sentinel-1 outperformed single-sensor approaches, aligning with the findings of Orynbaikyzy et al. (2020), Ofori-Ampofo et al. (2021), and Zoungrana et al. (2023).

Building on this multi-sensor approach, a new harmonized dataset i.e. Harmonized Landsat Sentinel-2 (HLS), which integrates Landsat-8's Operational Land Imager (OLI) and Sentinel-2's Multispectral Imager (MSI) has been developed (Claverie et al., 2018). HLS processing generates three key products: S10, which delivers MSI Surface Reflectance (SR) at 10m, 20m, and 60m resolutions; S30, a 30m MSI Nadir bidirectional reflectance distribution function (BRDF) - Adjusted Reflectance (NBAR); and L30, a 30m OLI NBAR - each derived from Level-1 input

data from Landsat-8 and Sentinel-2 (Claverie et al., 2018). Using HLS dataset Gao et al. (2023), proposed a method for near real-time detection of winter cover crop termination. Ju and Bohrer (2022) made use of NDVI time series derived from HLS to perform classification of wetland vegetation. Recent studies utilizing HLS have demonstrated promising results in improving crop classification accuracy (Berra et al., 2024), making it one of the ideal candidates for crop classification. Table 1 summaries the information on candidate satellite data and products that were described in the section.

Table 1. Summary of satellite data, including spatial resolution, number of bands, temporal revisits, satellites and launch date (data availability date in HLS case).

<i>Satellite/Data</i>	<i>Type</i>	<i>Spatial resolution</i>	<i>Data layer Or bands</i>	<i>Temporal revisit</i>	<i>Satellites</i>	<i>Launch date</i>
<i>Sentinel-1</i>	SAR	10 m	2 bands	12 days	Sentinel-1A Sentinel-1B Sentinel-1C	3 April 2014 25 April 2016 5 December 2024
<i>Sentinel-2</i>	Multispectral	10 m, 20 m and 60m	13 bands	10 days	Sentinel-2A Sentinel-2B Sentinel-2C	23 June 2015 7 March 2017 5 September 2024
<i>HLS</i>	Multispectral L30	30 m	10 bands and 5 ancillary bands (version 2)	16 days	Virtual constellation	Data available from 28 November 2015
	Multispectral S30	30 m	13 bands and 5 ancillary bands (version 2)	10 days	Virtual constellation	Data available from 11 April 2013

2.3 Machine learning and deep learning in crop type mapping

Machine learning (ML) and deep learning (DL) are subsets of artificial intelligence (AI). ML involves fitting mathematical models to observed data, while DL refers to a subset of ML that uses deep neural networks (Prince, 2023). Machine learning methods for image classification can be broadly categorized into supervised, unsupervised, and reinforcement learning approaches (Prince, 2023). Alami Machichi et al. (2023) further classify crop type mapping algorithms into three groups: (1) parametric ML, which assumes an underlying statistical distribution; (2) non-parametric ML, which does not impose any specific distribution assumptions; and (3) DL, which employs complex neural network architectures for classification.

Crop type classification using satellite data is a digital image classification task performed using both machine learning (ML) and, increasingly, deep learning (DL). Support Vector Machines (SVM), Random Forest (RF), eXtreme Gradient Boosting (XGB), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks are among the most widely used methods (Ma et al., 2019; Sheykhmousa et al., 2020; Alami Machichi et al., 2023; Victor et al., 2025). However, with the increasing spectral, temporal, and spatial resolution of satellite imagery,

ensemble-based models and DL techniques have outperformed classical models like SVM (Kussul et al., 2017; Alami Machichi et al., 2023). The success of deep learning can largely be attributed to the availability of vast amounts of training data, including both ground-truth field data and high-resolution satellite imagery (Victor et al., 2025).

Over the years, research has demonstrated the effectiveness of various DL architectures for crop type classification using time-series data. Models such as U-Net, Convolutional Neural Networks (CNNs), Recurrent CNNs (R-CNNs), Sparse Auto-Encoders (SAEs), and LSTM networks have been explored. Table 2 provides a summary of deep learning models, their reported accuracies, the satellite data they were trained on, and the number of crop classes they were used to differentiate. Notably, most DL models achieved high classification accuracies, often exceeding 95%.

Table 2. Summary of literature on crop type classification using deep learning models, satellite data used and reported accuracy.

<i>Literature</i>	<i>DL model</i>	<i>Crop type classes</i>	<i>Satellite data</i>	<i>Classification accuracy</i>
<i>Mazzia et al. (2020)</i>	RCNN	15	Sentinel-2	96.5%
<i>Alami Machichi et al. (2022)</i>	LSTM-Conv1D	4	Sentinel-2 +NDVI	95%
<i>Seydi et al. (2022)</i>	Dual Attention Convolutional Neural Network (DACNN)	6	Sentinel-2 +NDVI	95%
<i>Hamidi et al. (2024)</i>	SAE	7	RapidEye+ UAVSAR ¹	96-98%
<i>Dahiya et al. (2024)</i>	U-Net	2	Sentinel-2	95.6–98.4%

Recurrent Neural Networks (RNNs) are designed to effectively process sequential data. However, training RNNs presents challenges, particularly in capturing long-range dependencies, due to the vanishing gradient problem (Pascanu et al., 2013). LSTM networks, incorporating 1D convolution, have been shown to mitigate vanishing gradient issue and improve crop type classification accuracy (Alami Machichi et al., 2022). Similarly, Gated Recurrent Units (GRUs) were introduced to further enhance information retention. Sher et al. (2024) employed a dual-channel CNN-GRU feature fusion architecture, achieving an overall crop classification accuracy of 96.47% using Sentinel-2 and PlanetScope imagery, along with NDVI time-series data.

2.3.1 Transformer based model

Traditional CNN-based models, which are widely used for image classification, struggle to effectively capture crop texture at a 10-meter resolution, as seen in Sentinel-2 imagery (Geirhos et al., 2019). Moreover, deep learning models that rely on sequential data processing, such as RNNs

¹ <https://www.earthdata.nasa.gov/data/instruments/uavsar>

and LSTMs, require substantial computational resources, making them less practical for large-scale applications. These limitations highlight the need for models that can efficiently process both spectral and temporal information without excessive computational overhead.

Pixel-Set Encoder and Temporal Attention Encoder (PSETAE) was developed as a state-of-the-art model specifically for crop type classification. Garnot et al. (2020) designed PSETAE based on the transformer architecture (Vaswani et al., 2017) and the concept of pixel-set encoding (Qi et al., 2017). Unlike conventional deep learning models, PSETAE classifies crops by treating field parcels (field polygons) as objects, eliminating reliance on texture-based features, and effectively utilizing both spectral-spatial and temporal data (Garnot et al., 2020). One of the key advantages of PSETAE is its ability to process data in parallel, making it highly efficient for handling time-series satellite imagery. Garnot et al. (2020) initially developed the model for Sentinel-2 SITS. Ofori-Ampofo et al. (2021) later adapted it for Sentinel-1 and Sentinel-2 fusion datasets, introducing new data fusion methods to enhance classification accuracy. More recently, Weilandt et al. (2023) refined the model by implementing changes to improve temporal consistency.

In summary, the literature brings out significant progress that has been made in crop type classification using different satellite datasets, fusion methods and advanced ML and DL methods. However, key challenges remain, including limited availability of optical data during cloudy season, the need for higher spatial and temporal resolution, the need for integration of different satellite sensor data as data fusion techniques. While transformer-based models like PSETAE have enhanced classification accuracy and computational efficiency, their application to new geographies and datasets remains underexplored especially with respect to India. Building on this, our study is the first to apply PSETAE for crop type classification in southern part of India, utilizing Sentinel-1, Sentinel-2, and their fusion. We further extend the application of PSETAE to the newly available HLS dataset. This work of ours, highlights the adaptability and potential of PSETAE in addressing the challenges of crop classification in the Indian agricultural context.

3. Method and Materials

This section covers information related to study area followed by methodology. Methodology starts with how the ground truth data was collected, characteristics of the collected data, usage of GEE in data extraction and finally how PSETAE model works.

3.1 Study area

The study area is located in the state of Karnataka, Republic of India. Karnataka is one of the administrative states in southern India. Study area consists of two regions, marked by red rectangle as A and B in Figure 1. Region A includes Vijayapura district and parts of Yadgiri district, while Region B covers Dharwad, along with portions of Gadag, Belagavi, and Uttara Kannada districts. These regions were selected for their similar crops but slightly varied climates. The aerial distance between the two closest samples from these regions is approximately 100 km.

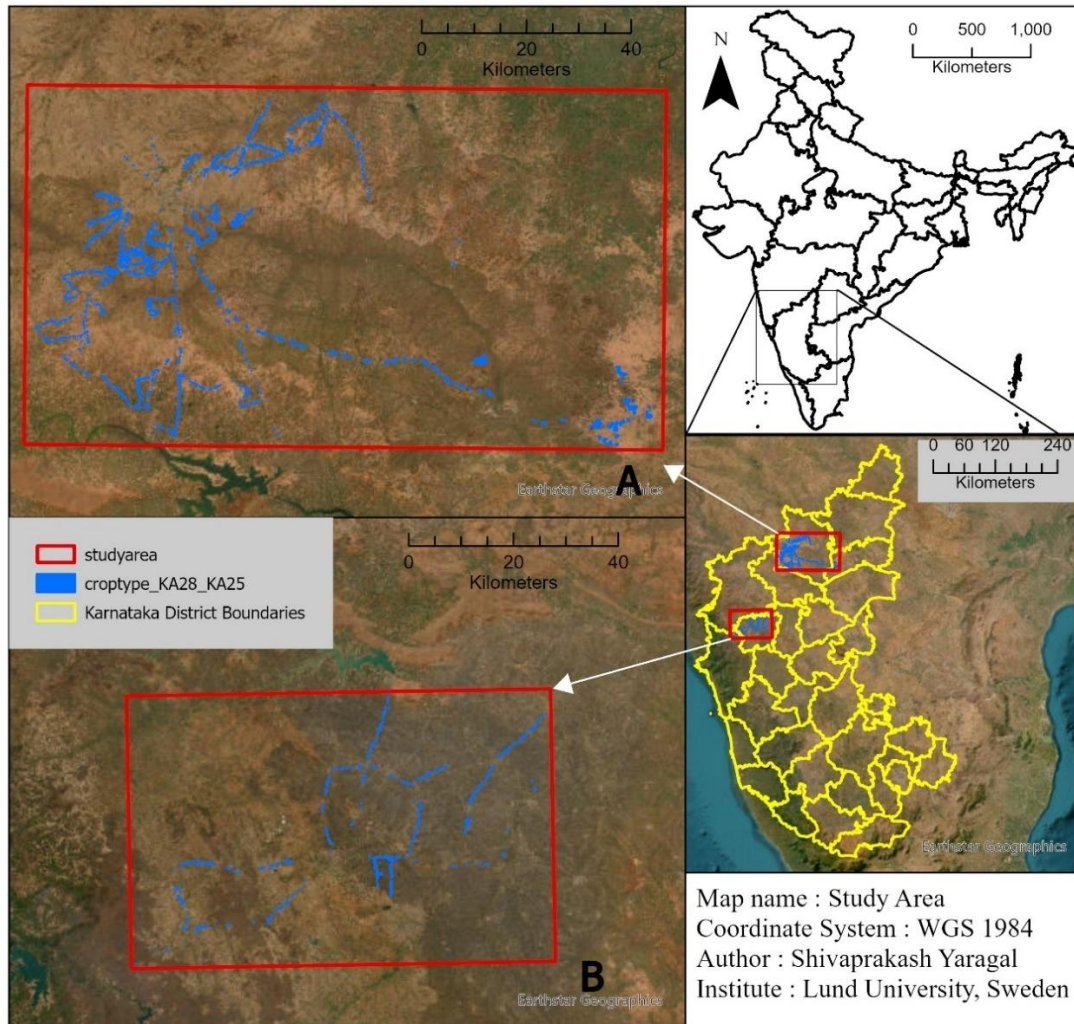


Figure 1. Study area in the State of Karnataka, India. Two regions, marked by red rectangles, were selected. Region A's data (Vijayapura and part of Yadgiri) was used for training, validation, and testing. Region B's data (Dharwad and parts of Gadag, Belagavi, and Uttara Kannada) was used for spatial generalization. Surveyed fields are represented using blue coloured polygons.

Region A has a semi-arid climate with hot summers, moderate monsoons, and mild winters, receiving 600–700 mm of annual rainfall, mainly from the southwest monsoon (Directorate of Census Operations, Karnataka, 2011). It falls under the northern dry agro-climatic zone (Ramachandra & Kumar, 2004), with agriculture being the main livelihood. Major crops include sorghum, pearl millet, pigeon pea, and chickpea, along with horticultural crops like pomegranate, lemon, grapes, and onion. Region B experiences a tropical wet and dry climate with 700–850 mm of annual rainfall. Summers are warm, and winters are mild (Directorate of Census Operations, Karnataka, 2011). It belongs to the northern transition zone (Ramachandra & Kumar, 2004), supporting crops such as sorghum, maize, chickpea, and horticultural produce like mango and banana. Irrigated areas also cultivate sugarcane and cotton. Some data from Haliyal area falls under the Hill Zone, which receives over 900 mm of rainfall (Ramachandra & Kumar, 2004). Both regions depend on the southwest and northeast monsoon winds for rainfall. Cropping in these regions takes place during the kharif season (June to September) and the rabi season (October to March of next year). The ground truth survey was conducted during the harvesting period of the rabi season, i.e., from January to March. Therefore, the collected data corresponds to rabi crops and was used for predicting crop types for the rabi season only.

3.2 Methodology

A workflow spread across three different software environments was implemented. Esri's products were used to conduct a mobile based ground truth survey, GEE environment was used for satellite data pre-processing and python environment was used to run open source and freely available PyTorch based deep learning model. Figure 2 gives an overview of the same.

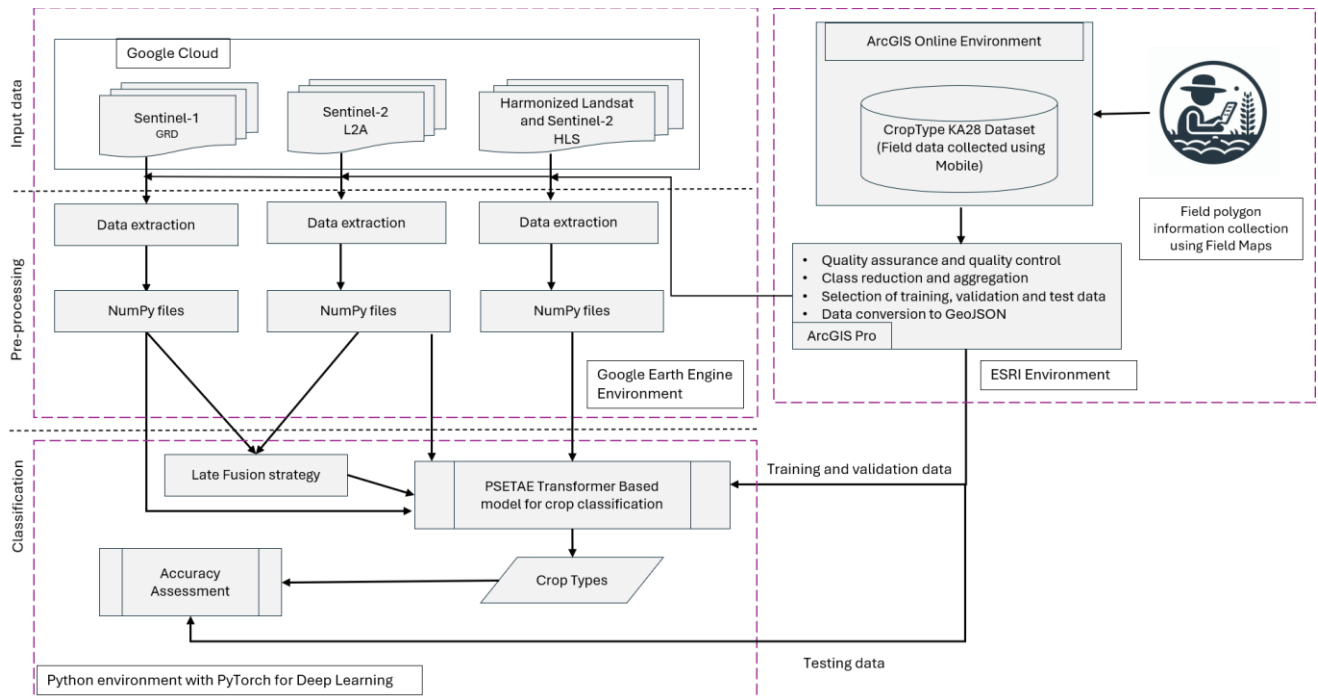


Figure 2. Workflow for the crop type classification showing unique integration of Esri's environment, GEE and Python environment.

Field data having information of polygons, crop type labels, and other ancillary data, was collected using a mobile-based application. This data was subsequently cleaned and processed using ArcGIS Pro. This involved quality checks, class reduction, and the selection of training, validation, and test samples. The validated field polygons were then used to extract SITS data from GEE. The extracted data was exported as NumPy files and stored locally for model training. Each dataset - Sentinel-1, Sentinel-2, and HLS - followed a distinct extraction method. Accordingly, data extraction was performed separately for each dataset within GEE. A PyTorch-based Python environment was set up to run the PSETAE model using these datasets as input.

For the fusion of Sentinel-1 and Sentinel-2 data, no separate preprocessing pipeline was implemented. Instead, fusion was performed internally within the PSETAE model during training. Model performance was evaluated using standard classification metrics, and this assessment was carried out individually for each dataset and fusion methods. In summary, field data collection and preprocessing were done using Esri tools, input satellite data was sourced, and data extraction took place in GEE, and model training and evaluation were conducted in a Python environment configured using PyTorch.

3.2.1 Mobile based data collection

For training crop classification algorithm, the most popular method is using ground truth data collected during field work using mobile based data collection methods (Alami Machichi et al., 2023). Data collected from such devices has produced appreciable crop classification results (Karale et al., 2014; Rahimi-Ajdadi & Khani, 2021; Koppaka & Moh, 2020, Choudhury et al. 2011; Wang et al., 2020). We leveraged this literature to design and configure a mobile application using ArcGIS Field Maps to capture field polygons, type of crop grown, capture photos and other related ancillary data.

3.2.2 Application design

Figure 3 illustrates the implementation of mobile based field surveying applications and monitoring dashboard using ESRI's ArcGIS Online (AGOL) product. AGOL is a software as a service (SaaS) product, functionally capable of creating data, hosting web feature service, storing images and application development.

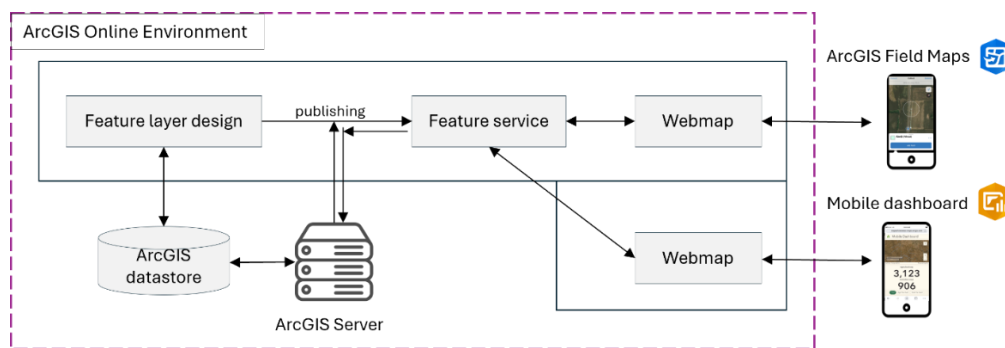


Figure 3. Architecture of a mobile-enabled geospatial data collection and visualization system configured using Esri's AGOL.

ArcGIS Datastore and ArcGIS server form the foundation of this implementation. ArcGIS datastore acts as a repository for spatial and attribute data. The schema of the feature layer was designed using a web interface. The designed layer was then published as a feature service with real-time editing and data synchronization capabilities. Implemented feature service has polygon as geometry and collects attribute related to crop type, date of polygon capture and photos of the field. The feature service was then integrated into a middle layer product called Webmaps which forms the configuration, visualization and editing middle layer. Webmaps were then used by ArcGIS Field Maps and also to design a mobile based dashboard for monitoring. Entire workflow is modular, scalable and cloud based. This helped us with easy data capture, proper planning, monitoring of the surveys, data backups and data dissemination. We used tutorial from Esri's website to configure and use these in our field work.

3.2.3 Survey planning

The survey was meticulously planned using geographical and crop type knowledge acquired from talking to local farmers. Over a period of 52 days (20 January 2024 to 19 March 2024), extensive fieldwork was conducted around more than 40 villages, covering over 2,000 kilometers by driving and 300+ kilometers on foot. This survey spanned across five districts, involving diverse terrains, cropping patterns, practices and different agricultural systems. Three sampling techniques were used for surveying - roadside transect sampling, L-shaped transect sampling and U-shaped transect sampling. Details of these methods are given in Appendix A. Field Maps and the monitoring dashboard were systematically used for planning and data collection. Implementation details are described in Appendix A. Empowered with surveying techniques, applications to collect and monitoring, a total of 4327 polygons were captured from Region A and Region B. After performing quality checks only 3941 polygons were deemed suitable for use in experiments. Figure 4 shows some of the fields visited during the planned survey.



Figure 4. Shows different type of crops in the field. (A) is Chickpea, (B) is onion, (C) is sorghum, (D) is grapes, (E) is wheat, (F) is safflower, (G) is rice, (H) is citrus, (I) is Pigeon pea, (K) is cotton, (L) is mango plantation containing sugarcane crop.

3.2.4 Field data characteristics

Mobile application was designed to collect data for 14 categories of agricultural crops and 10 categories of horticultural crops, in all 24 categories. However, 14 categories of crop data (both major and minor crops) were present during field surveys. Nine were agriculture crops and five were horticulture crops. From Region A, 3573 polygons were taken for processing. A crop type category called “Other crops” was introduced for keeping those agriculture crops which have numbers less than 0.5% of the total polygons in Region A. Based on this approach cotton, sunflower and safflower were merged to “Other crops” category. Citrus plantation (lemon), pomegranate and mango were also merged into one single category called fruit plantation. Effectively reducing horticultural crops into 3 categories and agricultural crops to 7. The final ten crop types were pigeon pea, chickpea, wheat, rice, sugarcane, sorghum, grapes, fruit plantation (including citrus plantation, pomegranate and mango), onion and other crops. Figure 5 shows the distribution of the data, which is characterized by imbalance in the dataset, with pigeon pea having the highest frequency, represented by 808 field polygons, followed by sugarcane with 536 polygons. For horticultural crops, grapes were the most frequently collected, with 394 polygons.

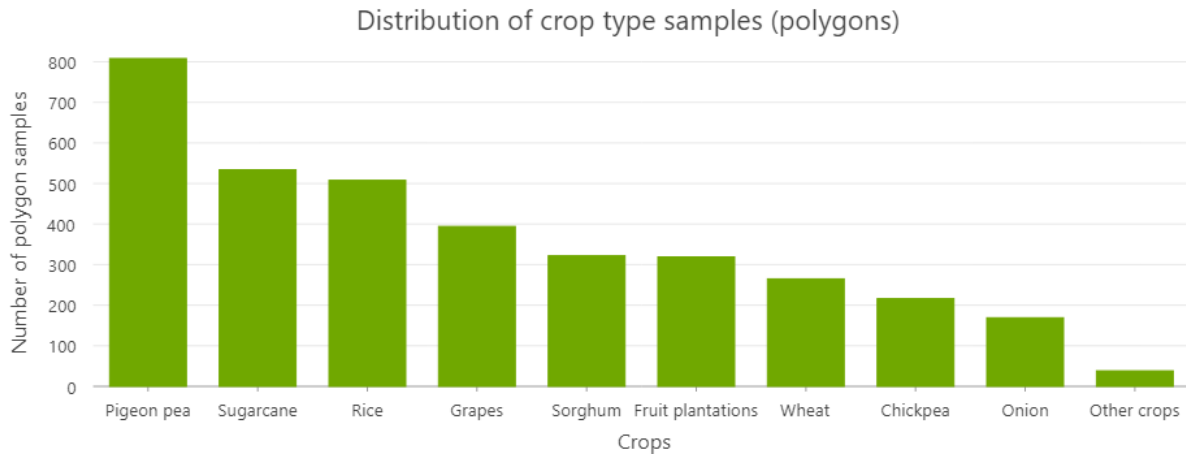


Figure 5. Distribution of surveyed field polygons after data cleaning and merging of the classes.

The average size of field polygons across all crops was 1.2 hectares. For agricultural crops, the average field size was 1.32 hectares, while for horticultural crops, it was 0.87 hectares. Among horticultural crops, grapes had an average field size of 0.9 hectares. Similarly, among agricultural crops, pigeon pea had the largest average field size of 1.88 hectares. Wheat fields had the smallest average size among agricultural crops, measuring around 0.7 hectares

Table 3 shows the various crop types, their average sizes in hectares, and the number of pixels based on the average field size. From the table, it is evident that pigeon pea, rice, sorghum, and chickpea fields contained more than 100 pixels per field polygon for the 10-meter resolution datasets, and more than 12 pixels for the 30-meter HLS dataset. Conversely, wheat, onion, and fruit plantations had fewer pixels per field polygon.

Table 3. Contains information on crop type, average size of polygons in hectares, number of 10-meter and 30-meter pixels that can fit in based on the average size.

<i>Crop type</i>	<i>Average size (hectares)</i>	<i>Number of 10-meter pixels</i>	<i>Number of 30-meter pixels</i>
<i>Pigeon pea</i>	1.88	188	20
<i>Sugarcane</i>	0.87	87	9
<i>Rice</i>	1.32	132	14
<i>Grapes</i>	0.96	96	10
<i>Sorghum</i>	1.23	123	13
<i>Fruit plantations</i>	0.58	58	6
<i>Wheat</i>	0.69	69	7
<i>Chickpea</i>	1.38	138	15
<i>Onion</i>	0.77	77	8

A total of 368 field polygons were selected from Region B to test spatial generalization. The dataset includes only five crops and was imbalanced. Figure 6 shows the data distribution, with chickpea being the most represented one and sorghum the least represented in Region B.

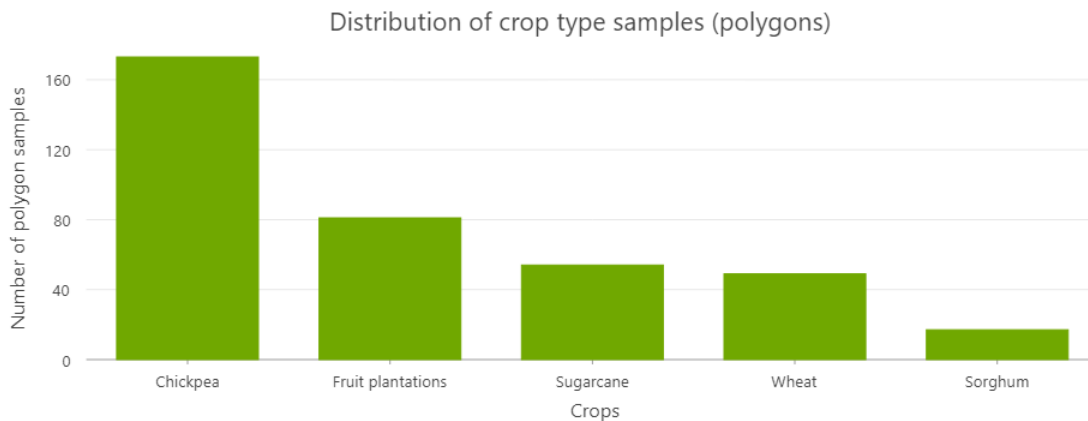


Figure 6. Distribution of surveyed field polygons in Region B.

3.3 Data processing using GEE

GEE is a cloud-based geospatial data processing platform that enables large-scale analysis of satellite imagery and geospatial datasets (Mandal et al., 2018). It provides access to a vast repository of Analysis Ready Data (ARD), a preprocessed dataset optimized for immediate use in applications such as crop type classification. For Sentinel-1, GEE uses ESA's Sentinel-1 toolbox workflow² to provide ARD products. GEE performs, border noise removal, thermal noise removal, radiometric correction, terrain correction, decibel conversion, multi-temporal speckle reduction (Zhao et al., 2018). For Sentinel-2, ESA's sen2cor processed Level-2A product was used. For extracting Sentinel-1 and Sentinel-2 data in the form of NumPy (.npz) files, a data method

² <https://developers.google.com/earth-engine/guides/sentinel1>

developed by Ofori-Ampofo et al. (2021) was used. Additionally, for the HLS dataset, both HLSS30 and HLSS30 were utilized. To extract the HLS dataset, a method similar to Ofori-Ampofo et al. (2021) was developed. Data extraction was performed by connecting to GEE using Python scripts. These scripts utilize the Python client library³ to authenticate, access, and interact with GEE, enabling the filtering, processing, and extraction of satellite data. The extracted data was then saved in the form of NumPy arrays for further analysis. The complete code implementation is publicly available in the GitHub⁴ (link provided below in footnote). For data extraction scripts, refer to the 2_data_extraction folder in the GitHub repository, and specifically the README_data_extraction.txt file for an overview of the data extraction scripts.

For the extraction of Sentinel-1, VH and VV polarization bands were used (Asam et al., 2022). For sentinel-2, all 10 bands except bands related to aerosol, water vapor and cirrus cloud bands were used which is in line with experiments conducted by Ofori-Ampofo et al. (2021) and Asam et al. (2022). Similarly, in the case of the HLS datasets (HLSS30 and HLSS30), the bands that contributed little to vegetation analysis -such as the coastal band, cirrus band, panchromatic and water vapor bands- were excluded. A detailed list of the spectral bands used in the study is provided in Table 4. The temporal window for data extraction for all satellite datasets spanned from September 2024 to March 2025 a total of 7 months (210 days). This period is in line with rabi season with added one extra month of September 2024.

Data augmentation using band ratios and vegetation indices such as NDVI was avoided in this study to maintain a uniform comparison across different satellite datasets. Relying on indices can obscure the contribution of raw spectral bands. By excluding such indices, the analysis remains grounded in the original spectral data, ensuring that performance differences across datasets reflect the characteristics of the raw inputs rather than the influence of derived features.

Table 4. Bands extracted from satellite dataset for training the model.

<i>Satellite/Data</i>	<i>Bands extracted</i>	<i>Band excluded</i>	<i>Additional info</i>
<i>Sentinel-1</i>	VV,VH	HH,HV	
<i>Sentinel-2</i>	B2,B3,B4,B5, B6,B7,B8,B8A,B11,B12	Aerosol (B1), Water vapor (B9) and SWIR Cirrus (B10)	
<i>HLSS30</i>	B2, B3, B4, B5, B6, B7, B10, B11	Costal(B1), Cirrus (B9)	B8 i.e. panchromatic doesn't exist in HLSS30
<i>HLSS30</i>	B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12	Aerosol (B1), Water vapor (B9) and SWIR Cirrus (B10)	

³ <https://developers.google.com/earth-engine/guides>

⁴ https://github.com/shivaprakash23/cropclassification_karnataka_psetae/

3.4 PSETAE model

This section discusses the PSETAE architecture that classifies the medium resolution SITS data. Field polygon serves as the basic unit of data extraction. A satellite observation of a given field polygon is denoted by spectro-spatial and temporal parameters, represented as a multidimensional tensor of shape (T, B, H, W) - where T represents the number of temporal observations, B the number of spectral channels or bands, and H and W the dimensions in pixels of the bounding box that contains the field (Garnot et al., 2020). The primary task of PSETAE is to learn from multiple bands (B) across T time-series satellite observations for each field polygon. For each field polygon, the model randomly samples S pixels that lie within the boundary to represent the field. These S pixels are sampled from within the boundary of the polygon to capture its spectral characteristics.

3.4.1 Pixel-Set Encoder

The Pixel-Set Encoder (PSE) processes the spatial and spectral dimensions through a shared Multi-Layer Perceptron (MLP). This MLP transforms S pixel values into higher-dimensional feature representations. The transformed features are then pooled by calculating mean and standard deviation along the pixel axis, effectively averaging the learned parameters across all sampled pixels. This statistical pooling creates a permutation-invariant representation that handles the variable number of pixels across different fields. The PSE concatenates the pooled spectral features with geometric information about the field polygon. This adds spatial context to the spectral information. This combined spectro-spatial information is then processed through another MLP, resulting in an embedding tensor represented as "e" in the architecture (Figure 7). It contains rich information about both the spectral characteristics and geometric properties of the field.

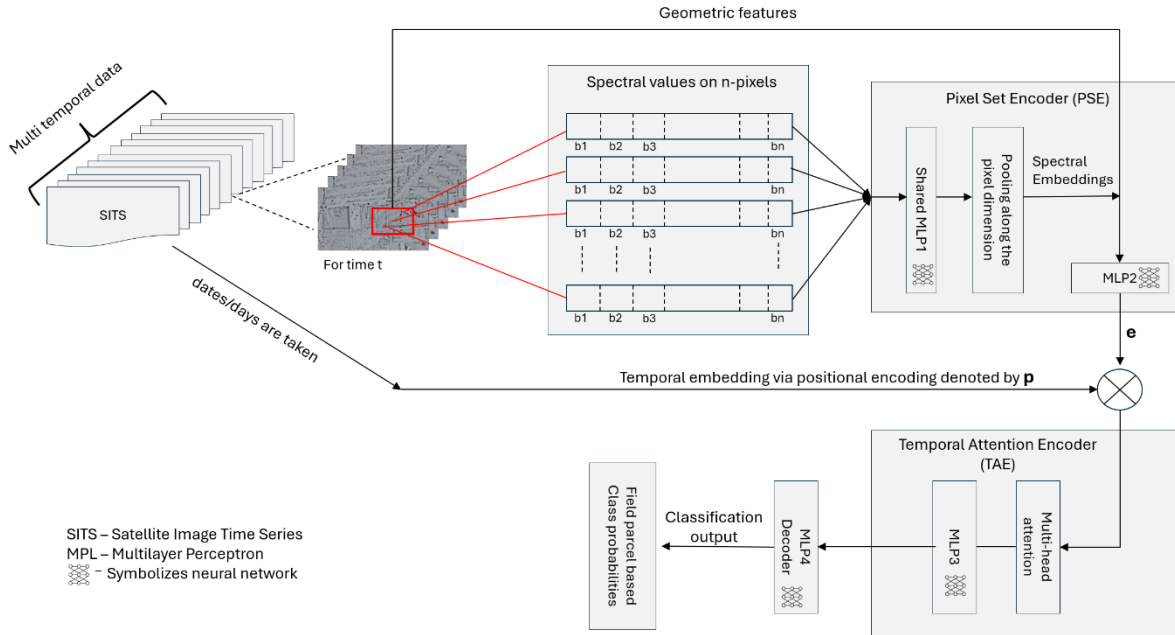


Figure 7. PSETAE architecture showing Pixel-set encoder and Temporal attention encoder components.

3.4.2 Temporal Attention Encoder

Temporal Attention Encoder (TAE) is responsible for identifying and focusing on the most informative time steps in the satellite image time series. The temporal information needed for TAE is incorporated into the output of PSE by a positional encoder. For each value of "e" at time t for a polygon, a temporal position encoding "p" is added as shown in the provided Figure 7. The implementation uses sinusoidal functions for this positional encoding, with a parameter T controlling the period of these functions. As t changes over the time series, both "e" and "p" change, capturing the phenological evolution of the field. The combined embedding (e+p) is then processed by the TAE, which implements a transformer-based attention mechanism inspired by Vaswani et al. (2017).

TAE has a modified self-attention approach which learns to focus on important time-steps that would be most helpful in detecting the crop type, similar to identifying the optimal month when crop differentiation becomes most apparent due to phenological differences. The outputs from all attention heads are concatenated and processed through another MLP3. This MLP3 summarizes the attention mechanism, transforming the attention-weighted representations into a final temporal embedding. The final component is a classification MLP4 decoder that takes the output from the TAE and produces class probabilities.

3.5 Implementation

Garnot et al. (2020)'s PSETAE was adapted to multi-sensor data fusion by Ofori-Ampofo et al. (2021). GitHub implementation of Ofori-Ampofo et al. (2021) has been used for single sensor and multi-sensor implementations using Sentinel-1 (hereafter referred to as S1) and Sentinel-2 (hereafter referred to as S2) dataset. A similar method was developed and adapted for processing HLSS30 and HLSS30 dataset. The complete code implementation is publicly available in GitHub and link has been provided in section 3.3.

3.5.1 Data extraction

For data extraction, cloud cover was kept below 20% for S2, HLSS30 and HLSS30. The extracted data were stored as NumPy files in "DATA" folder and related metadata were stored in "META" folder. Each field polygon produced one NumPy file in DATA folder and its related metadata like date got stored in date.json, labels in label.json and if any polygon was ignored, they were stored in ignored.json files.

3.5.2 Hyperparameter tuning

We divided the data of Region A into training, validation and test dataset following 60:20:20 ratio. Accordingly, we got 2144 parcels for training, 714 for validation and 715 for testing. Data extracted using GEE were then used to train the PSETAE model. All the experiments were conducted using NVIDIA GeForce RTX 4060 GPU. The following are hyperparameters that were kept constant.

- Architectural hyperparameters
- Length of maximum sequence for positional encoding (lms) – 210 for 7 months
- Total time period T – 366
- Number of attention heads – 4

We performed hyperparameter tuning for each dataset type using the search space defined in the “Search Space” column of Table 5. Each experiment was run for 20 epochs. The results are summarized in Table 5. The optimal learning rate was found to be 0.0001 for both S1 and S2, and 0.001 for the HLS datasets. The best-performing n-pixel values were 64 for S1, 128 for S2, and 16 for both HLS datasets. The optimal dropout rates were 0.3 for S1, 0.2 for S2, and 0.1 for both HLS datasets. Finally, the optimal minimum sampling parameters were 16 for S1, 42 for S2, and 8 for both HLS datasets.

Table 5. Hyperparameters and search space for S1, S2, HLS datasets.

Hyperparameters/sensors	Search space	S1	S2	HLSS30	HLSL30
learning rate	(0.01, 0.001, 0.0001)	0.0001	0.0001	0.001	0.001
n-pixel	(16,32,64,128)	64	128	16	16
batch size	(32,64)	32	32	64	64
drop out	(0.1, 0.2, 0.3)	0.3	0.2	0.1	0.1
Minimum sampling	(4,8) for HLS (8,16) for S1 (24,32,42) for S2	16	42	8	8

3.5.3 Experimental design

Once hyperparameters were determined, four types of experiments were conducted. The first type involved evaluating individual datasets: S1, S2, HLSS30, and HLSL30. The second type tested three fusion methods combining S1 and S2 data. The third type examined model performance during the initial growing period i.e. September to December (Refer section *Cloud cover statistic* of Appendix C for more information on selecting this period), to determine if fusion methods could outperform individual sensors (S1 and S2) in terms of accuracy. In each of these experiments, five sub-experiments with the same hyperparameters were conducted, varying only the random seed. This approach helped us to compute the average of the evaluation metrics used to assess the model performance. In all the experiments training, validation and test data were kept the same to maintain consistency in the experiments. Finally, the fourth experiment involved applying the best-performing models - developed during the first and second experiments using S1, S2, HLSS30, HLSL30, and the three fusion methods - for inference using unseen data from Region B.

3.5.4 Choice of fusion methods

We followed the fusion methods as described by Ofori-Ampofo et al. (2021). These were implemented using S1 and S2 datasets. We performed two layer level fusion and one decision level fusion. Figure 8 shows the PSE Fusion, TAE Fusion - both representing layer level fusion

method- and Statistical Average Fusion (as decision level fusion). A brief explanation of each is given below.

PSE Fusion : S1 and S2 data are first processed independently through two separate PSE modules, which extract features specific to S1 and S2. The outputs are then temporally aligned and concatenated along the feature dimension, followed by a shared TAE and then classification. This approach preserves S1 and S2 features while allowing their interaction in the temporal domain. PSE Fusion is robust when S1 and S2 provide complementary information.

TAE Fusion : S1 and S2 data are processed independently through their own PSE and TAE allowing each stream to learn S1 and S2 specific spatial-spectral and temporal learning. The outputs are concatenated after TAE. This maximizes S1 and S2 specific learning, allowing each stream to capture unique temporal patterns. TAE Fusion excels when S1 and S2 have distinct temporal signatures.

Statistical Average Fusion : In this method, S1 and S2 are processed through entirely separate pipelines up to the prediction stage. The final class probabilities from each stream are averaged to produce the final output. This is an ensemble approach that leverages the strengths of S1 and S2's independent prediction. Decision-level fusion like Statistical Average Fusion is easy to implement and is interpretable. For Statistical Average Fusion, we used simple averaging (as suggested Ofori-Ampofo et al. (2021)) of class probabilities as shown in Equation 1.

$$p_{a_i}^{S1S2}(x) = \frac{p_{a_i}^{S1}(x) + p_{a_i}^{S2}(x)}{2} \quad (1)$$

Where $p_{a_i}^{S1}(x)$ and $p_{a_i}^{S2}(x)$ are the class probabilities output for S1 and S2 results.

PSE Fusion, performed after PSE module requires temporal alignment of S1 and S2 datasets. In the current study area, over an entire period of seven months (September 2024 to March 2025), S1 has 16 observations, while S2 has 24 observations (cloud cover < 20%). To enable temporal concatenation, the number of S2 observations were reduced to 16 by selecting, for each S1 date, the closest corresponding S2 observation. This was done using the nearest-neighbor matching in the time domain. This temporal alignment step is specific to the PSE Fusion approach and is not applied in the other two fusion methods.

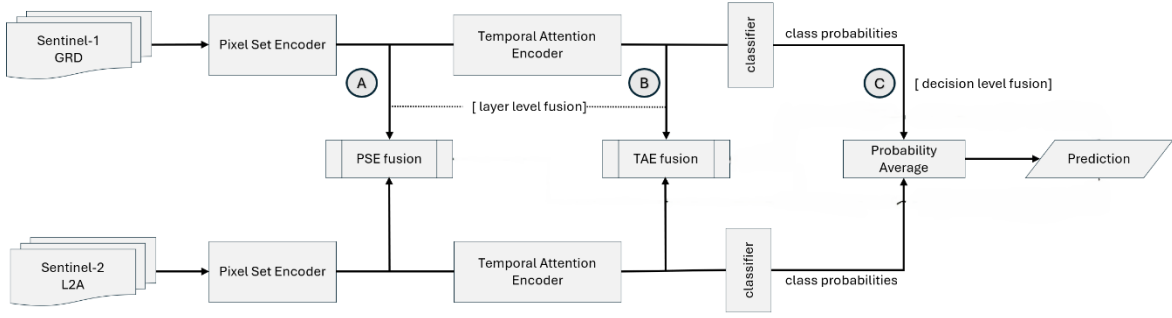


Figure 8. Fusion methods highlighting fusion after PSE module (A), after TAE module (B) and finally at decision level (C). Source: Ofori-Ampofo et al. (2021).

3.5.5 Evaluation metrics

To assess the performance of the classification model, we used Accuracy, Macro Intersection over Union (Macro IoU), Macro F1 Score, and Cohen's Kappa Value. These metrics provide a comprehensive evaluation of the model. Accuracy (Equation 2) is the proportion of correctly classified samples out of the total number of samples. It gives an overall measure of a model's performance, showing how often the classifier is correct (Rainio, Teuho, & Klén, 2024). Accuracy varies between 0 to 100 and is expressed in percentage. The higher the accuracy, the better the model is.

$$Accuracy = \frac{\sum_{i=1}^K TP_i}{\sum_i (TP_i + FP_i + FN_i + TN_i)} \quad (2)$$

IoU, given by Equation 3, measures the overlap between the predicted and actual classes for each class individually (Rainio, Teuho, & Klén, 2024). Macro IoU is the average IoU across all classes given by Equation 4. Micro IoU varies between 0 to 1 and is expressed as decimal number. Higher IoU signifies good classification.

$$IoU_i = \frac{TP_i}{(TP_i + FP_i + FN_i)} \quad (3)$$

$$Macro IoU = \frac{1}{K} \sum_i IoU_i \quad (4)$$

Precision (Equation 5) measures how many of the predicted positive instances are actually correct (Rainio, Teuho, & Klén, 2024). Recall (Equation 6) measures how many of the actual positive instances were correctly predicted. F1 Score (Equation 7) is the harmonic mean of Precision and Recall (Rainio, Teuho, & Klén, 2024). It balances precision and recall. Macro F1 Score, given by Equation 8, is the arithmetic mean of the F1 Scores computed independently for each class (Opitz, 2024). Macro F1 score varies from 0 to 1 and is expressed as decimal number. Macro F1 score is

for models and F1 score is for specific class. Higher Macro F1 score for models signifies better classification. Additionally, a higher F1 score signifies better classification performance for that crop compared to other crops with lower F1 scores.

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \quad (5)$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (6)$$

$$F1\ score_i = \frac{2 * Precision_i * Recall_i}{Precision_i + Recall_i} \quad (7)$$

$$Macro\ F1 = \frac{1}{K} \sum_{i=1}^K F1\ score_i \quad (8)$$

Where

TP_i is True Positives for class i

FP_i is False Positives for class i

FN_i is False Negatives for class i

TN_i is True Negatives for class i

K is Number of classes

Cohen's Kappa (as given in Equation 9) is a metric that measures how well a classification model performs compared to random guessing (Opitz, 2024). Kappa values have a range of 0 to 1. A Kappa value of 1 indicates perfect agreement, whereas a value of 0 means the model performs no better than random guessing. The higher Kappa value signifies better agreement between the predicted and actual classifications, indicating that the model is performing well above the level of random chance which is 0.

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (9)$$

Where

p_o is observed agreement

p_e is expected agreement by chance.

κ is Cohen's Kappa

4. Results

In this section, we first examine how model's learning varies with respect to different datasets and fusion methods. This is followed by comparative analysis using the evaluation metrics discussed in section 3.5.5. Finally, we assess spatial generalization capabilities by applying the inferencing on unseen dataset collected from Region B.

4.1 Evaluation of model learning

S2 achieves the highest validation accuracy reaching up to 96.13%, outperforming S1 and all the three fusion methods. Both HLSS30 (with 92.26% validation accuracy) and HLSSL30 (with 91.28% validation accuracy) deliver strong validation accuracy, with HLSS30 slightly outperforming HLSSL30. S1 shows the lowest performance among all datasets with validation accuracy reaching up to 86.87% . Among the fusion experiments, S1+S2 TAE Fusion performs the best with validation accuracy of 95.23% which was lower than S2.

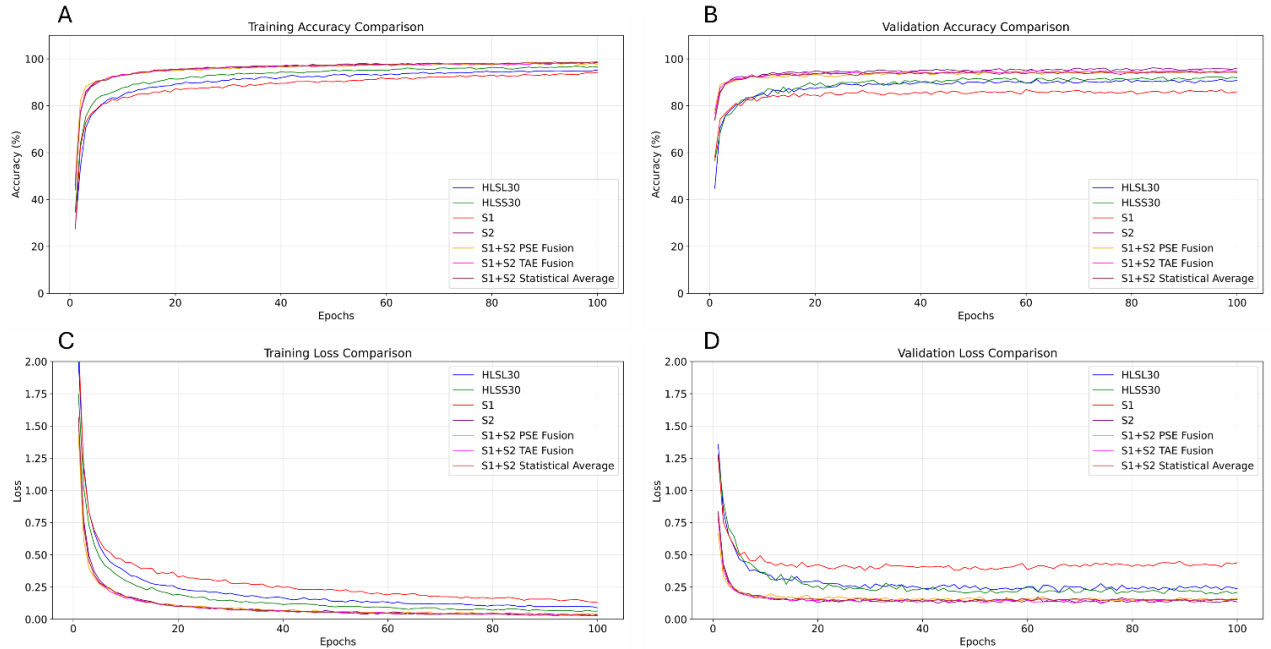


Figure 9. Training accuracy, validation accuracy, training loss and validation loss across 7 experiments over 100 epochs: training accuracy (image A), validation accuracy (image B), training loss (image C) and validation loss (image D).

S2 and the fusion methods converge faster than S1 and HLS sensors, reaching near-optimal performance within fewer epochs as seen from image D i.e. validation loss graph of Figure 9. S2 achieved the lowest validation loss (0.121). The fusion methods showed comparable validation loss values (ranging from 0.125 to 0.139), while S1 exhibited the highest validation loss (0.379). Both the HLS datasets showed the intermediate lowest validation loss, with 0.208 for HLSSL30 and 0.195 for HLSS30 - which is higher than S2 and fusion methods, but lower than S1. From image B, all methods show steep improvement in the first 10-15 epochs, with S2 and fusion methods

achieving more than 90% validation accuracy within just 5 epochs. In contrast, S1 shows a slower convergence pattern, requiring approximately 30 epochs to stabilize at 85%.

On overfitting tendency, S1 exhibits the largest gap between training and validation accuracy (5~7%), indicating a higher degree of overfitting compared to others. S2 maintains a relatively small training-validation gap (~2-3%). In the fusion strategy, Statistical Average Fusion (here on S1+S2 Stat Avg Fusion) shows the smallest training-validation gap among fusion methods (~2-3%). HLSS30 and HLSS30 show moderate training-validation gaps (~4-6%), which fall between S1 and S2. All methods showed significantly diminished improvement rates after epoch 50, with approximate minimal gains (<0.5% accuracy) in later epochs. On validation fluctuations, S1 showed the highest amplitude of fluctuations in validation accuracy throughout training, indicating less stable learning dynamics. In contrast, S2 and S1+S2 Fusion methods showed remarkably smooth validation accuracy curves after epoch 30, suggesting more consistent convergence. These trends were indicative of the positive role played by dropout-based regularization. Specifically, dropout rates of 0.3 (S1), 0.2 (S2), and 0.1 (HLS datasets) helped constrain overfitting and stabilize the training process across different sensor types.

4.2 Comparative analysis using evaluation metrics

The quantitative evaluation of different datasets and fusion methods revealed significant performance variations across all metrics. Table 6 summarizes the results of 5 runs for each of the dataset and fusion methods. S2 consistently outperformed all other approaches, achieving the highest test scores across all metrics (95.97% accuracy, 0.87 Macro IoU, 0.92 Macro F1, and 0.95 Kappa). Surprisingly, S1 demonstrated the lowest test accuracy of 87.16% with a substantial performance gap compared to S2, with an accuracy difference of 8.81%. All three fusion methods (S1+S2 PSE Fusion, S1+S2 TAE Fusion, and S1+S2 Stat Avg Fusion) yielded comparable results, with S1+S2 PSE Fusion slightly outperforming the others.

Table 6. Results of experiments with evaluation metrics.

<i>Sensor/Metrics</i>	<i>Accuracy</i>	<i>Macro IoU</i>	<i>Macro F1 score</i>	<i>Kappa</i>
<i>S1 only</i>	87.16	0.69	0.79	0.85
<i>S2 only</i>	95.97	0.87	0.92	0.95
<i>S1+S2 PSE Fusion</i>	94.73	0.84	0.90	0.94
<i>S1+S2 TAE Fusion</i>	94.03	0.82	0.89	0.93
<i>S1+S2 Statistical Average</i>	94.48	0.83	0.89	0.94
<i>HLSS30</i>	93.06	0.80	0.87	0.92
<i>HLSS30</i>	91.08	0.74	0.82	0.90

The fusion methods showed marked improvement over S1 alone (S1+S2 PSE Fusion improves accuracy by 7.57%) but fell short of S2's performance by a small margin (S1+S2 PSE Fusion trails S2 by 1.24% in accuracy). Among the HLS products, HLSS30 (93.06% accuracy) outperformed

HL30 (91.08% accuracy). HLS products outperformed S1 but fell short of S2 by roughly 4% for HL30 and by roughly 3% for HLSS30. The Kappa coefficient, which accounts for the possibility of agreement occurring by chance, follows a similar pattern (of accuracy) with S2 achieving the highest value (0.95), followed by fusion methods (0.93-0.94), HLSS30 (0.92), HL30 (0.90), and S1 (0.85).

On Macro F1 score and Macro IoU, S2 stands out with the highest Macro IoU (0.87) and Macro F1 score (0.92). Among the fusion methods, S1+S2 PSE Fusion achieved the strongest results (Macro IoU: 0.84, Macro F1: 0.90), closely followed by S1+S2 Stat Avg Fusion and S1+S2 TAE Fusion. All fusion methods outperformed S1 and HLS products. The S1 recorded the lowest scores (Macro IoU: 0.69, Macro F1: 0.79). The HLS products demonstrated intermediate performance, with HLSS30 (Macro IoU: 0.80, Macro F1: 0.87) consistently outperforming HL30 (Macro IoU: 0.74, Macro F1: 0.82).

To understand how metrics vary across five runs, box plots were created showcasing variance for S1, S2, HLS datasets and fusion methods. Figure 10 shows these box plots. The highest variance in the accuracy was observed for S1+S2 Stat Avg Fusion method, while the lowest was noted for S2. Both the HLS datasets show the lower variance compared to S1 and fusion methods. Among fusion methods, S1+S2 Stat Avg Fusion showed higher variance compared to S1+S2 PSE Fusion and S1+S2 TAE Fusion. Image A of Figure 10 provides a visual summary of these observations.

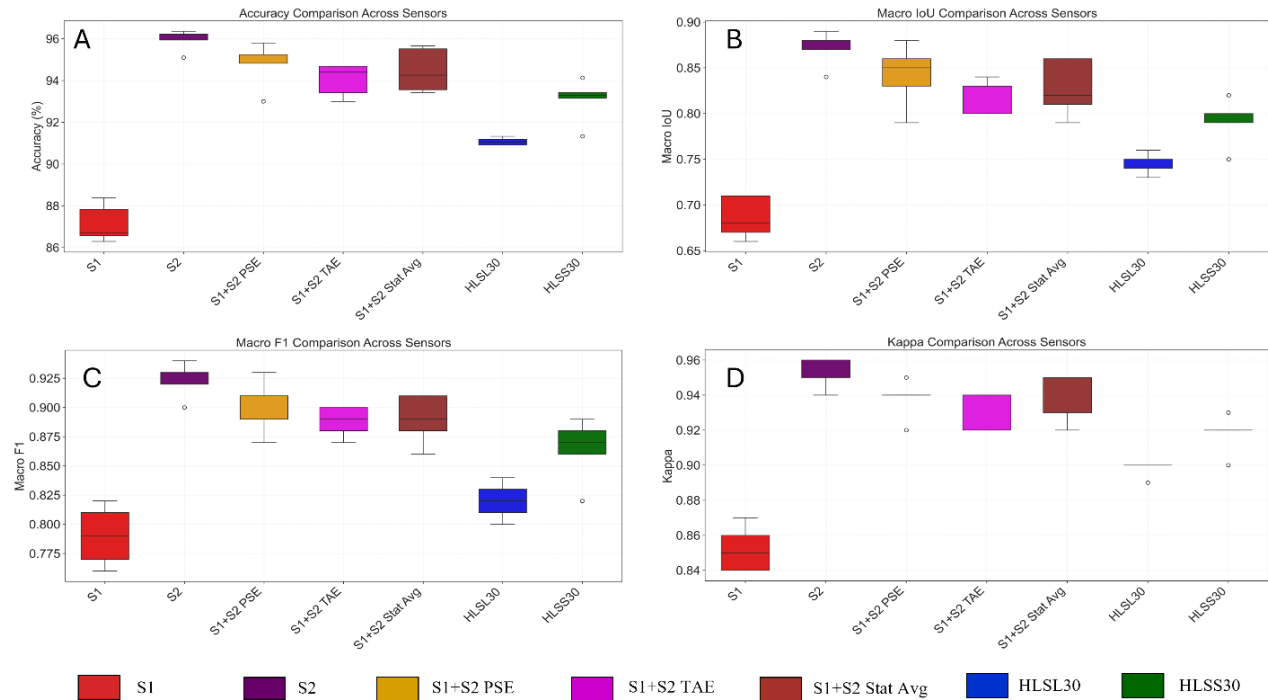


Figure 10. Boxplots of evaluation metrics Accuracy, Macro IoU, Macro F1 and kappa over five runs.

A similar pattern of variance was observed for Macro IoU, as shown in image B of Figure 10. For Macro F1, all datasets and fusion methods showed variance, with the highest being for S1 and the

lowest being for S2. HLSS30 and HL30 showed moderate variance. The kappa score graph (image D) revealed that output for HLS datasets were consistent, indicated by narrow box. Among fusion methods, S1+S2 PSE Fusion showed higher consistency when compared to S1+S2 TAE Fusion and S1+S2 Stat Avg Fusion. Overall, the variances observed in the fusion methods were greater than those for the HLS datasets.

Table 7 shows detailed class wise F1 score which were averaged over five runs for all experiments. The class-wise F1 scores reveal distinct performance patterns across different dataset and fusion methods. S2 consistently achieved the highest F1 scores across all crop classes averaging 0.92, significantly outperforming S1 (average F1: 0.79). The fusion methods demonstrated performance comparable to S2, with S1+S2 PSE Fusion showing the best results among other fusion methods. The performance gap is particularly pronounced for challenging classes like "Other crops" where S2 (0.65) outperformed S1 (0.20). The HLS datasets demonstrated an interesting intermediate performance. HLSS30 consistently outperformed HL30 across all crop classes, with particularly notable differences for chickpea (HLSS30: 0.90, HL30: 0.82) and wheat (HLSS30: 0.83, HL30: 0.75). Overall, HLSS30 achieved performance closer to the fusion methods, while HL30 positions between S1 and S2.

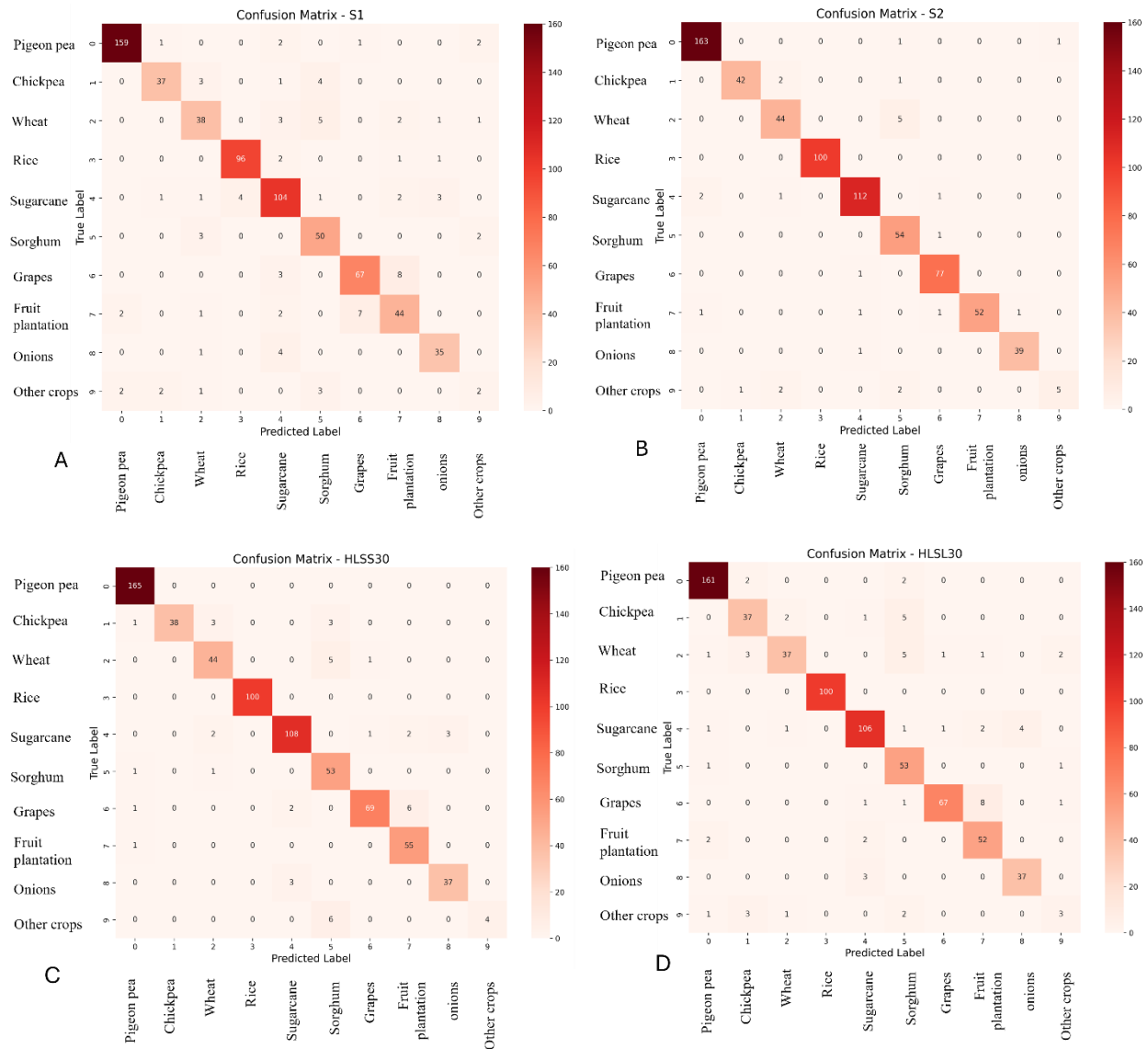
Table 7: Class wise F1 scores average over five runs.

<i>Crop type class</i>	<i>Crop name</i>	<i>S1 only</i>	<i>S2 only</i>	<i>S1+S2 PSE Fusion</i>	<i>S1+S2 TAE Fusion</i>	<i>S1+S2 Statistical Average</i>	<i>HLSS30</i>	<i>HL30</i>
0	Pigeon pea	0.96	0.98	0.98	0.98	0.98	0.98	0.98
1	Chickpea	0.85	0.94	0.94	0.92	0.90	0.90	0.82
2	Wheat	0.74	0.90	0.85	0.83	0.85	0.83	0.75
3	Rice	0.95	1.00	1.00	1.00	1.00	0.99	1.00
4	Sugarcane	0.86	0.96	0.95	0.93	0.95	0.93	0.93
5	Sorghum	0.81	0.92	0.89	0.89	0.89	0.86	0.84
6	Grapes	0.87	0.97	0.97	0.98	0.97	0.95	0.93
7	Fruit Plantations	0.79	0.96	0.95	0.95	0.94	0.92	0.89
8	Onion	0.85	0.96	0.91	0.90	0.92	0.91	0.90
9	Other crops	0.20	0.65	0.60	0.49	0.49	0.39	0.17

Pigeon pea and rice emerged as the most accurately classified crops across all satellite datasets and fusion methods. Rice achieved near-perfect classification (F1: 0.95-1.00). Similarly, pigeon pea maintained consistently high F1 scores (0.96-0.98) across all configurations. The fusion methods maintained the high performance for these classes as well. For grapes and fruit plantations, S2 (0.97, 0.96) substantially outperformed S1 (0.87, 0.79), with fusion methods maintaining performance comparable to S2. This pattern is consistent across other crops like sugarcane, sorghum, and onion, where S2 provides F1 scores 10-11 percentage points higher than S1. The fusion methods closely tracked S2's performance for these classes, with S1+S2 PSE Fusion

generally achieving the highest F1 scores among fusion methods. S1+S2 TAE Fusion outperformed S2 only in case of grapes.

Figure 11 shows confusion matrices for the best-performing run of each of the seven experiments: S1, S2, HLSS30, HLSSL30, S1+S2 PSE Fusion, S1+S2 TAE Fusion, and S1+S2 Stat Avg Fusion. Confusion matrix are the tools to better understand the confusion and separability among the crops. Most confusion occurred for S1 followed by HLS dataset. S2 and fusion methods were able to resolve the confusion in efficient manner.



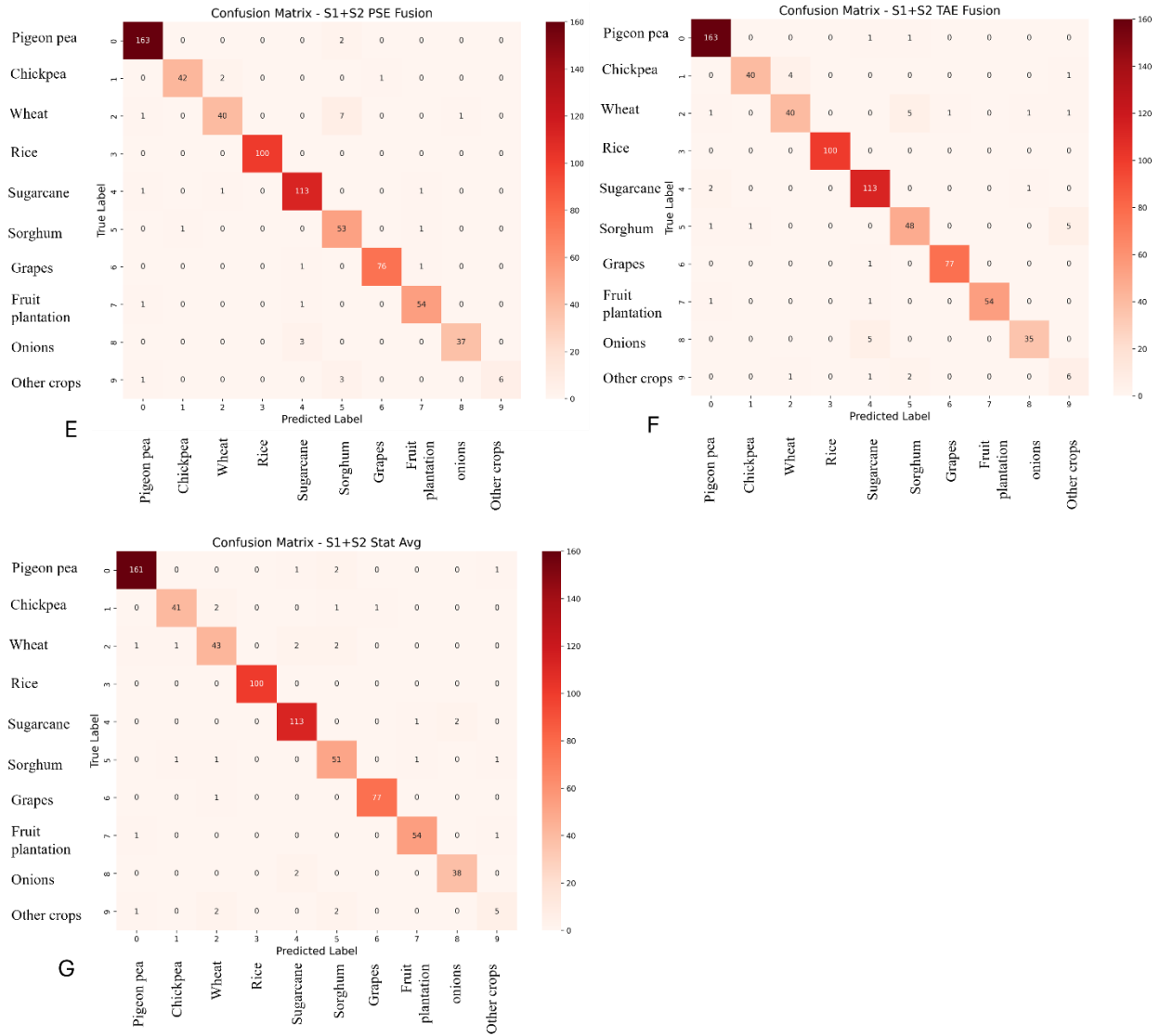


Figure 11. Confusion matrices for the best-performing run of each of the seven experiments: S1 (image A), S2 (image B), HLSS30 (image C), HLSSL30 (image D), S1+S2 PSE Fusion (image E), S1+S2 TAE Fusion (image F), and S1+S2 Statistical Averaging Fusion (image G).

Among agricultural crops, the highest confusion was observed between sorghum, chickpea, and sugarcane. For horticultural crops, misclassification frequently occurred between plantation and grape fields, particularly with the S1 and HLS datasets. However, S2 and fusion methods were able to distinguish these classes more effectively, resulting in improved class separability.

The model using the S2 dataset that achieved the highest accuracy was used to develop the crop type classification map for the test dataset. Figure 12 shows the predicted classes of the field polygons for the Region A test dataset. Images L, M, N, and O are zoomed-in views of selected areas. Green dots on the map indicate correct class predictions, while red dots represent incorrect predictions. To generate a complete map of the region, digitized field polygons from every village

in the area would be required. In our study, we had access only to our own collected data from Region A; therefore, the map represents only the collected dataset.

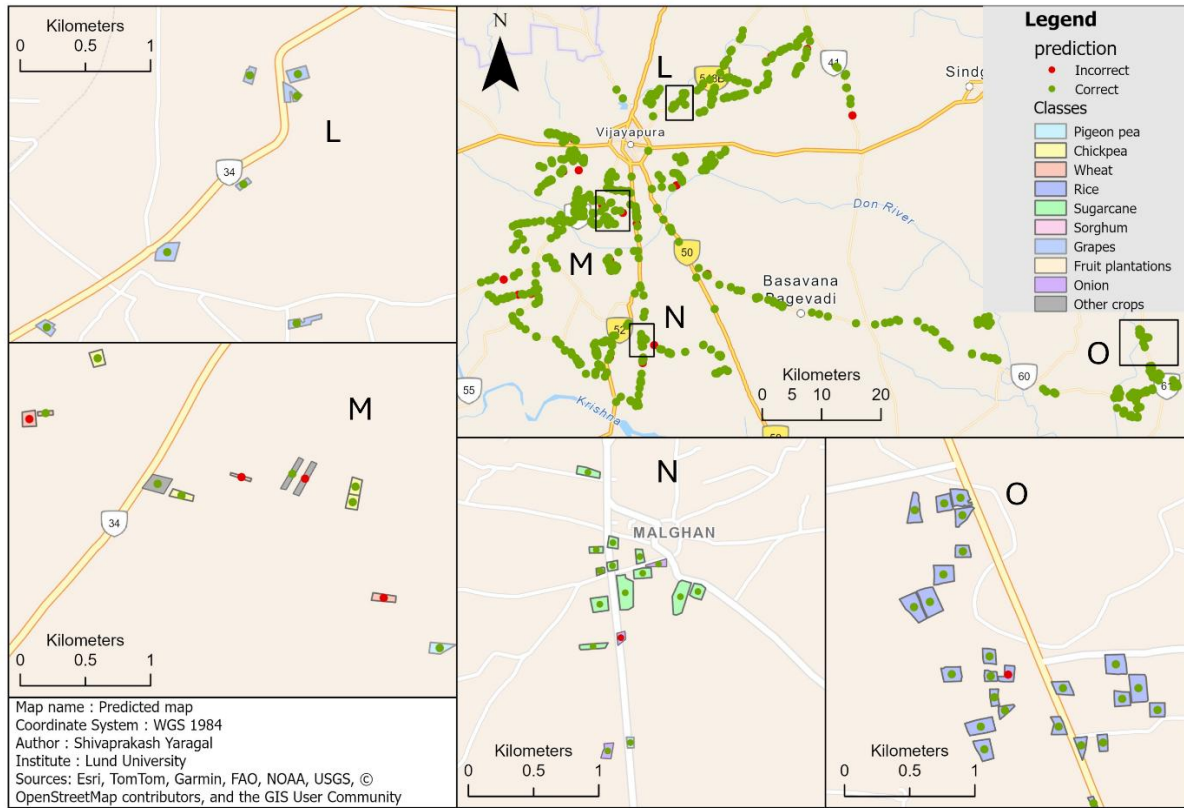


Figure 12. Predicted map containing four zoomed in areas L, M, N, O. Image L is from area dominated by grapes. Image M has mixed crops. Image N is dominated by sugarcane and onion. Image O is rice growing area. Green dots are the correct prediction, and red dots are the incorrect prediction.

4.3 Performance during initial growing period

The results of the third type of experiments, which evaluated the accuracy of S1, S2 and fusion of both during the months of September to December, are presented in Table 8. S1 achieved an accuracy of 76.56% while S2 achieved an accuracy of 88.54%. Among the fusion methods, the S1+S2 TAE Fusion method resulted in the highest accuracy at 90.36%, followed by the S1+S2 PSE Fusion at 89.15%, and the S1+S2 Stat Avg Fusion at 88.93%. All these fusion methods outperformed both S1 and S2 for this period. The improvement of the fusion methods over S1 ranges from 12.37% (S1+S2 Stat Avg Fusion) to 13.80% (S1+S2 TAE Fusion). Compared to S2, the fusion methods offer an increase in accuracy ranging from 0.39% (S1+S2 Stat Avg Fusion) to 1.82% (S1+S2 TAE Fusion). Thus, during initial growing period, the fusion methods provided an accuracy improvement in the range of approximately 12–14% over S1 and 0.4–1.8% over S2.

Table 8. Model performance during the cloud-affected period (September–December).

Sensor/Metrics	Accuracy
<i>S1 only</i>	76.56
<i>S2 only</i>	88.54
<i>S1+S2 PSE Fusion</i>	89.15
<i>S1+S2 TAE Fusion</i>	90.36
<i>S1+S2 Statistical Average</i>	88.93

4.4 Model inferencing on Region B

Table 9 tabulate the results of inferencing on Region B data, highlighting accuracy metrics with respect to datasets and fusion methods. The cross-region (Region B) inference results reveal a reversal in performance compared to the in-region (Region A) testing, offering valuable insights into model generalization . The most remarkable finding is the significant performance inversion between SAR and optical sensors. S1 (69.11%) substantially outperformed S2 (50.27%) in the new region, completely reversing the pattern observed in the original test area where S2 was the top performer (95.97% vs. S1's 85.95%). The fusion methods demonstrated good resilience in cross-region inferencing, with S1+S2 PSE Fusion (68.85%) nearly matching S1's performance. The relative performance of fusion methods (S1+S2 PSE Fusion > S1+S2 Stat Avg Fusion> S1+S2 TAE Fusion) differed from the in-region testing. The HLS dataset showed particularly poor cross-region generalization, with HLSS30 achieving only 41.46% accuracy. Interestingly, HLSSL30 (52.17%) outperformed HLSS30 despite showing lower performance in the original region,

Table 9. Results for Region B with Accuracy as metrics and accuracy at given confidence level.

<i>Sensor/Metrics</i>	<i>Accuracy</i>	<i>Accuracy at a Given Confidence (>80%)</i>
<i>S1 only</i>	69.11	79.32
<i>S2 only</i>	50.27	61.39
<i>S1+S2 PSE Fusion</i>	68.85	88.57
<i>S1+S2 TAE Fusion</i>	62.84	73.90
<i>S1+S2 Statistical Average (Stat Avg)</i>	67.76	84.39
<i>HLSS30</i>	41.46	51.04
<i>HLSSL30</i>	52.17	78.70

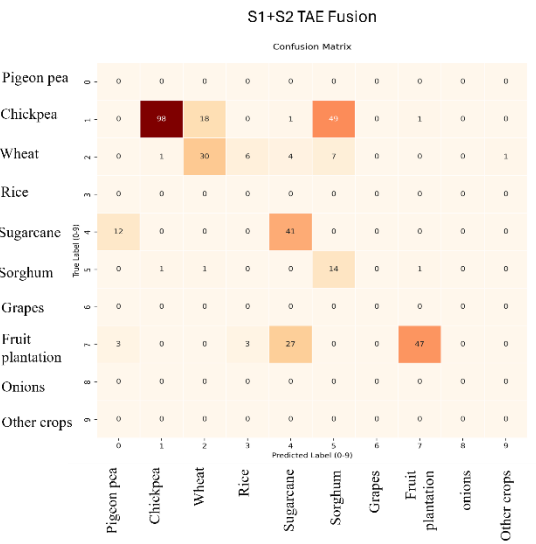
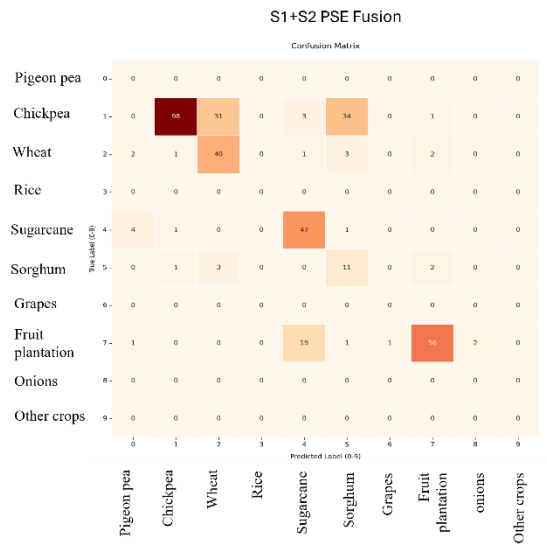
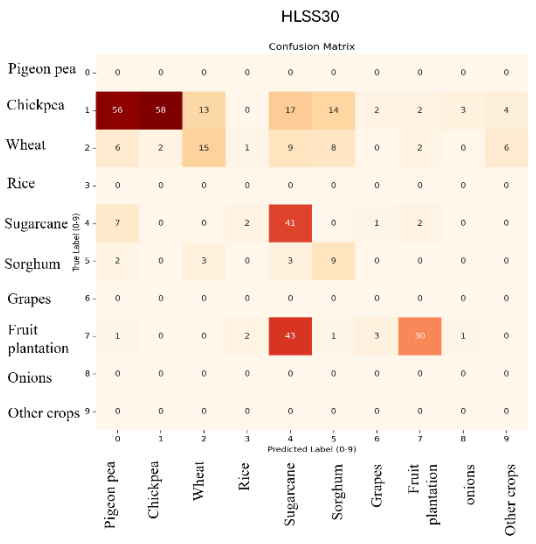
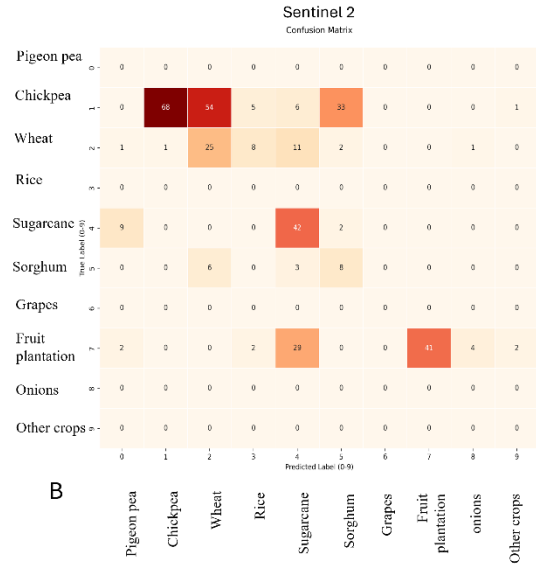
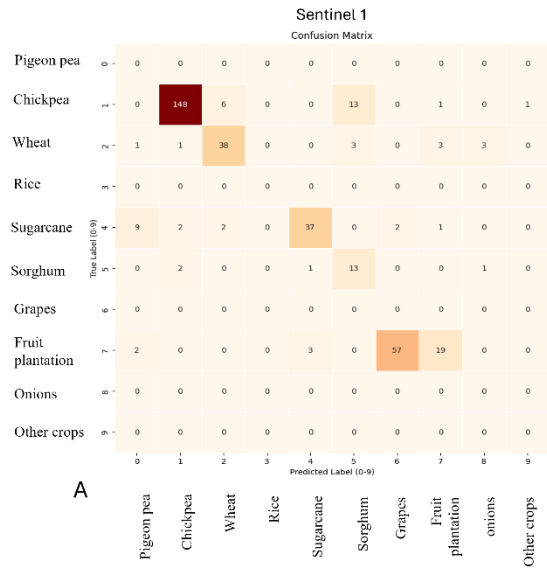
To further assess the reliability of the model predictions, we computed the "Accuracy at a Given Confidence" metric by considering only those predictions where the model's confidence exceeded 80%. The results, summarized in Table 9, demonstrate a range of selective accuracies across S1,S2, and HLS datasets and fusion methods. Analysis reveals that fusion methods demonstrated the highest selective accuracies, with S1+S2 PSE Fusion achieving 88.57%, followed by S1+S2 Stat Avg Fusion (84.39%) and S1+S2 TAE Fusion (73.90%). S1 showed remarkable selective accuracy (79.32%), substantially outperforming S2 (61.39%). The HLS products exhibited the most divergent behavior, with HLSSL30 (78.70%) performing well above HLSS30 (51.04%).

The class-wise F1 score (Table 10) evaluation on Region B data bring out some interesting information. S1 exhibited good performance in identifying chickpea (0.92), wheat (0.80), and sugarcane (0.79), while S2 recorded notably low values for these classes (0.58, 0.37, and 0.58 respectively). However, S2 outperformed S1 in classifying fruit plantations (S2=0.68 vs S1=0.36). Among the fusion methods, S1+S2 Stat Avg Fusion achieved the best balanced performance, when compared to other two fusion methods. It performed particularly well in classifying fruit plantations (0.83) surpassing both S1 and S2. It also maintained good performance for sugarcane (0.78) comparable to that of S1. The S1+S2 PSE Fusion method demonstrated robustness with fruit plantations (0.79) and sugarcane (0.76). Notably, all fusion methods outperformed S1 and S2 in classifying fruit plantations.

Table 10. Class wise F1 scores Region B.

<i>Crop type class</i>	<i>Crop name</i>	<i>S1</i>	<i>S2</i>	<i>S1+S2 PSE Fusion</i>	<i>S1+S2 TAE Fusion</i>	<i>S1+S2 Statistical Average</i>	<i>HLSS30</i>	<i>HLSL30</i>
0	Pigeon pea	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	Chickpea	0.92	0.58	0.73	0.73	0.71	0.50	0.78
2	Wheat	0.80	0.37	0.65	0.31	0.51	0.38	0.13
3	Rice	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	Sugarcane	0.79	0.58	0.76	0.65	0.78	0.49	0.12
5	Sorghum	0.57	0.26	0.33	0.32	0.28	0.37	0.06
6	Grapes	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	Fruit Plantations	0.36	0.68	0.79	0.72	0.83	0.51	0.82
8	Onion	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	Other crops	0.00	0.00	0.00	0.00	0.00	0.00	0.00

The results of HL30 shows good performance in classifying chickpea (0.78) and fruit plantations (0.82), while HLSS30 maintained poor performance across multiple classes similar to that of S2. HL30 performed poorly with wheat and sugarcane reaching only (0.13 and 0.12). Across all models, sorghum remained a challenging class, with the highest F1 score reaching only 0.57 (by S1). S1 performed seemly good for three out of five classes and S2 performed poorly on all accounts. Fusion methods along with HL30 consistently performed well for fruit plantations. In resolving the confusion related to mixed cropping (Chickpea and Sorghum), S1 performed better than S2, fusion methods and HLS datasets, resulting in higher accuracy for cross-region generalization. S2 performed the poorest. S1 struggled to differentiate between grapes (which were not there) and fruit plantation in horticulture crops and chick pea and sorghum in agriculture crops. Significant confusion was observed in S2's result especially for chick pea, wheat and fruit plantation. A similar confusion was noted in HL30's output as well, especially in chick pea, wheat and sugarcane. The fusion methods, however, demonstrated improved class separability compared to S2 alone. Figure 13 shows the confusion matrices for predictions on Region B.



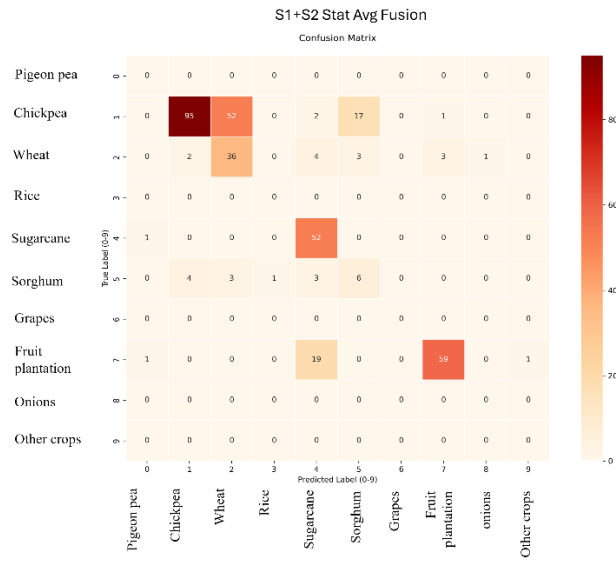


Figure 13 : Confusion matrices for crop classification using Region B data across all seven experimental setups: S1 (image A), S2 (image B), HLSS30 (image C), HLSL30 (image D), S1+S2 PSE Fusion (image E), S1+S2 TAE Fusion (image F), and S1+S2 Statistical Averaging Fusion (image G).

5. Discussion

5.1 Interpretation of the results

In this study, we presented a workflow that integrates field surveys, different satellite data, various data fusion methods, and transformer-based classification algorithm to classify the ten main crops in a diverse Indian farming setup. The use of data fusion methods, a transformer-based deep learning model for crop type classification, and the incorporation of the HLS dataset were unique characteristics of this work, making it one of a kind. While other crop classification efforts (Teluguntla et al., 2015; Gumma et al., 2016; Mandal et al., 2018; Wang et al., 2020; Rao et al., 2021; Vijayan et al., 2023; Nagendram et al., 2023) have primarily focused on major agricultural crops like rice, wheat, millets, and sorghum, our study is the first to attempt classification of horticultural crops such as grapes and fruit plantations, along with other agricultural crops. Moreover, while previous studies did not include spatial generalization experiments, we designed our inferencing process on unseen data to assess spatial generalization accuracy. This approach has been implemented for the first time in the Indian farming setup and is a unique aspect of this study.

The outcome of this study is in agreement with the existing literature (Hamidi et al., 2024; Rao et al., 2021; Ofori-Ampofo et al., 2021; Song et al., 2021), showing that S2 (hereafter referred to as Sentinel-2) outperformed the S1 (hereafter referred to as Sentinel-1) by up to 8-9% in accuracy metrics. Across the entire period (September 2024 to March 2025), the fusion of Sentinel-1 and Sentinel-2 did not substantially improve accuracy compared to Sentinel-2. At best, the results were on par with Sentinel-2. This is in contrast with existing literature (Ofori-Ampofo et al., 2021; Blickensdörfer et al., 2022).

During the early cropping period (September 2024 to December 2024), all fusion methods outperformed both Sentinel-1 and Sentinel-2, with S1+S2 TAE Fusion achieving the highest accuracy. This, we believe, is due to Sentinel-2 losing data quality because of cloud cover. It is also spectrally difficult to separate crops during this early growth stages, especially with medium-resolution pixels. Cloud-related data filtering also introduces temporal inconsistency in Sentinel-2 observations, weakening its ability to capture crop dynamics during early growth stages. On the other hand, Sentinel-1, unaffected by cloud cover, captures structural information that helps in learning phenological patterns early in the season. Fusion methods leverage the strengths of Sentinel-1 to compensate for the inconsistencies of Sentinel-2, resulting in higher accuracy. S1+S2 TAE Fusion, a good example, achieved this by maintaining separate PSE-TAE processing stream for both Sentinel-1 and Sentinel-2, effectively avoiding the entanglement of noisy Sentinel-2 data and thereby improving overall accuracy.

Across the entire period (September 2024 to March 2025), HLS dataset demonstrated accuracy more than Sentinel-1 and slightly lower, though comparable to Sentinel-2 and fusion methods,

marking a significant finding for crop mapping in the Indian subcontinent. This is aligned with the results reported by Parreiras et al. (2022) and McCormick et al. (2025). Among HLS products, HLSS30 outperformed HL30, indicating that this may be due to its higher number of spectral bands. Both HLS datasets achieved good accuracy for all nine crop types present in the study area. This is a substantial finding, especially considering the use of 30-meter medium-resolution satellite data to generate crop-level maps at a subcontinental scale.

5.1.1 Class specific interpretation

The class-wise F1 score analysis revealed interesting information related to crop identifiability across satellite datasets and fusion methods. Pigeon pea and rice emerged as the most accurately classified crops across all satellite datasets (F1: 0.95-1.00), suggesting their distinctive signatures in both optical and SAR imagery. Results of rice were in line with finding of Mandal et al. (2018) Vijayan et al. (2023) and Nagendram et al. (2023). Pigeon pea and rice were the least confused classes owing to their highly monoculture crop practice in the region. Most of the omissions and commissions, leading to misclassification, were observed among sorghum, wheat, and chickpea (as seen in confusion matrix in Figure 11). These confusions stem from the complex cropping practices in the region. Sorghum, for instance, is a versatile crop used both as fodder and food, and is often intercropped with cash crops like chickpea - either as single strands, entire rows or along field edges. Similarly, wheat is grown as a subsistence crop and as a replacement crop in areas affected by waterlogging. We suggest that such cropping pattern and crop mixing further contributed to spectral confusion. The field data that we collected also confirmed the presence of such crop mixing. More details on the observed crop mixing is given in the Appendix B.

Onion showed some confusion with sugarcane, which can be attributed to the local cropping patterns - areas where sugarcane is harvested early (in November or December) are often quickly replanted with summer onions for the next three to four months. This cropping cycle may have introduced temporal confusion between the two classes. Another interesting case of confusion was also observed between grapes and fruit plantations, particularly in the results of S1, HL30, and HLSS30. This confusion aligns with findings from a study conducted in Germany by Asam et al. (2022), which reported similar confusion between vineyard crops and other fruit plantation. Fusion methods were able to resolve this confusion making better use of both Sentinel-1's structural features and Sentinel-2's spectral features. Although Sentinel-2 independently was able to achieve similar F1 score. The "Other crops" class exhibited both high commission and omission, primarily due to the inclusion of three mixed crops - cotton, safflower, and sunflower - which introduced both spectral and temporal confusion. This is one class where the fusion methods were able to consistently produce reliable results. However, the performance for this class was also limited by the very small number of available samples.

5.2 Implications and limitations

The results of spatial generalization have significant implications for operational crop mapping systems. The cross-regional (Region B) performance of Sentinel-1 and fusion methods highlights the valuable contribution of SAR data to model accuracy, emphasizing the importance of including such data when developing models for wide-area applications. The dramatic performance differences between in-region (Region A) and cross-region testing (Region B) (e.g., Sentinel-2: 95.97% vs. 50.27%) underscore the critical need for comprehensive validation across diverse geographies before operational deployment. For applications that demand both high accuracy and geographic transferability, the S1+S2 PSE Fusion approach demonstrates the most balanced performance, effectively leveraging the complementary strengths of optical and SAR sensors to achieve robust classification across varied regional conditions. The improved performance of fusion approaches suggests that combining Sentinel-1's SAR data with Sentinel-2's spectral information helps mitigate the limitations of individual sensors.

F1 scores from the spatial generalization experiments (Table 10) offer a glimmer of hope for operational use. Notable performances include Sentinel-1's results for chickpea (0.92), wheat (0.80), and sugarcane (0.79); S1+S2 Statistical Average Fusion's result for fruit plantations (0.83); and HLSL30's result for fruit plantations (0.82). These values are quite good and comparable to some of the results from Region A. Such outcomes suggest the potential of combining predictions from different data sources – such as Sentinel-1, Sentinel-2 and HLS – during inference to improve the accuracy in operational crop classification.

The study also reveals an important hierarchy pattern in crop classification performance. In the primary study region i.e. Region A, Sentinel-2 consistently outperformed all approaches (96% accuracy), while Sentinel-1 demonstrated the lowest performance (87% accuracy). However, this hierarchy dramatically reversed in cross-region testing, where Sentinel-1 (69% accuracy) substantially outperformed Sentinel-2 (50% accuracy). This remarkable inversion suggests that SAR data captures more transferable information about crops that remains consistent across geographical contexts, while optical data is more sensitive to regional variations in spectral characteristics due to differences in soil conditions, farming practices, and phenological timing. This finding challenges the common assumption that optical sensors provide superior crop classification in operational setting.

Fusion approaches demonstrated remarkable resilience across different testing scenarios, suggesting their value as stability enhancers in operational crop monitoring systems. While fusion methods could not surpass Sentinel-2's performance in the primary region (falling short by 1-2% in accuracy), they significantly outperformed Sentinel-1 by 7-8%. Most notably, in cross-region application, S1+S2 PSE fusion (69% accuracy) nearly matched Sentinel-1's performance while substantially outperforming Sentinel-2. This adaptability indicates that a well designed fusion approaches can utilize Sentinel-2's spectral richness for local precision while incorporating Sentinel-1's transferable features for geographic resilience.

Classification confusions in Region A, became more pronounced during the generalization test. Notably, Sentinel-2 showed significant confusion among chickpea, sorghum, and wheat, underscoring its limited generalization capability. This limitation of Sentinel-2 was also evident in the HLSS30 confusion matrix, which showed a similar pattern. We suggest that this is likely due to the Sentinel-2 contribution in the harmonized product. In contrast, Sentinel-1 handled these distinctions more effectively. Fusion methods leveraged this strength of Sentinel-1's to compensate for Sentinel-2's weaknesses, thereby mitigating the confusions.

An interesting observation was the shift in confusion patterns across regions. In Region A, confusion was evident between fruit plantations and grapes. However, in Region B - where grapes were absent - the confusion shifted to fruit plantations and sugarcane. The confusion is a result of observed crop mixing in the Region B where sugarcane was present in different ways within mango plantation (refer Appendix B for more details). The shift in confusion appears to be primarily influenced by Sentinel-2, supporting our understanding that the confusion is more spectral in nature rather than structural.

The analysis of training dynamics revealed important patterns with implications for deploying the model. Sentinel-2 and fusion methods demonstrated remarkably faster convergence compared to Sentinel-1 and HLS sensors, reaching near-optimal performance within just 15 epochs. This rapid convergence suggests potential for significant computational savings, as training could be shortened by up to 50% without substantial performance loss. The HLS products demonstrated an interesting intermediate performance, positioning them as potential compromise solutions. HLSS30 consistently outperformed HL30 across all metrics in the Region A (93% vs. 91% accuracy), likely due to its higher number of bands and better preservation of Sentinel-2's spectral characteristics. However, this relationship also reversed in cross-region testing, where HL30 (52% accuracy) outperformed HLSS30 (41% accuracy), suggesting that Landsat-derived features may offer better generalization for cross-regional applications. This can be seen with respect to HL30's F1 score for fruit plantations (0.82). The overall performance gradient from Sentinel-1 to HL30 to HLSS30 to Sentinel-2 in Region A provides a spectrum of options for practitioners to balance performance against other considerations such as data availability, processing requirements, and application scope.

The analysis of "Accuracy at a Given Confidence" (>80%) provides crucial insights into model reliability in operational scenarios. This selective accuracy metric demonstrated that Sentinel-1 and fusion methods maintained substantially higher performance when considering only high-confidence predictions, particularly in cross-region application. While Sentinel-2's overall cross-region accuracy was only 50%, its selective accuracy rose to 61%, still significantly lower than S1 (79%) and S1+S2 PSE Fusion (89%). This finding has direct implications for deployment strategies, suggesting that incorporating confidence thresholds could substantially improve reliability, especially when using fusion approaches. Furthermore, the exceptional performance of

fusion methods under high-confidence conditions (S1+S2 PSE Fusion achieving 89% selective accuracy) indicates their potential for creating robust classification systems based on confidence level.

While this study presents several interesting results, it is essential to acknowledge its limitations. Firstly, the analysis was limited to rabi season, thereby restricting its temporal generalizability. Secondly, the training dataset was relatively small in comparison to those utilized in similar studies (Wang et al., 2020; Ofori-Ampofo et al., 2021; Asam et al., 2022). Additionally, the study focused only on major crops observed during the field visit, which may not fully represent the true diversity of crops present in the study area. The research also did not incorporate ancillary data such as soil moisture, soil type, irrigation information, or vegetation indices, all of which could enhance crop type classification. At the fusion level, the study was limited to existing layer-level and decision-level fusion methods, and did not explore band-level fusion or advanced decision-level fusion techniques. These limitations highlight key research gaps and thus pave the way for future studies aimed at developing more comprehensive and generalizable crop classification models.

6. Conclusion and Future Research

This thesis presents a comprehensive workflow for crop type classification using mobile based field surveys, multi-sensor satellite data (Sentinel-1, Sentinel-2, and HLS), fusion methods, and transformer-based deep learning model. We mapped and classified ten major crops across a heterogeneous agricultural region in and around Vijayapura, State of Karnataka, India.

The following conclusions were drawn from the study

1. Sentinel-2 consistently achieved the highest in-region classification accuracy compared to Sentinel-1 and fusion methods.
2. Fusion methods, particularly the PSE Fusion, provided stable accuracy - higher than Sentinel-1's and slightly lower than Sentinel-2's accuracy.
3. The Statistical Average Fusion, despite being simpler than PSE Fusion, also emerged as a strong alternative.
4. All fusion methods outperformed the HLSS30 and HL30 datasets in terms of classification accuracy.
5. In terms of spatial generalization, Sentinel-1 and fusion methods performed better than Sentinel-2 and HLS datasets.

The findings of this thesis have several implications. First, the inversion of classification accuracy across regions challenges the common practice of reporting only in-region metrics and highlights the need for cross-region validation to assess true operational capability. Second, SAR and SAR-based fusion methods appear more robust for operational systems, particularly in geographically diverse areas. Third, the improvement in accuracy when filtering predictions by confidence indicates that estimating model confidence could be an important component of future crop classification systems.

Future research could extend this work in several directions. Spatially, PSETAE should be tested across a broader range of agro ecological zones - including mountainous terrains, rainforest-adjacent areas, and arid regions - to assess adaptability. Temporally, studies should investigate performance over multiple agricultural seasons and years. A critical next step is to explore crop classification during the kharif season (June to September), when heavy cloud cover makes SAR-optical fusion particularly valuable for Indian agriculture. At architecture level, introducing phenology-aware or density-aware sampling strategies may further improve classification accuracy of fusion methods. Finally, advanced decision-level fusion of Sentinel-1, Sentinel-2, and HLS dataset should be explored to refine predictions and produce more accurate agricultural insights.

In conclusion, this study offers a solid foundation for developing scalable and transferable crop classification datasets and methods using satellite data. By examining the differences between in-region and cross-regional accuracy, this study paves the way for development of location-specific geo-adapted crop type classification models.

References

- Alami Machichi, M., El Mansouri, L., Imani, Y., Bourja, O., Lahlou, O., Zennayi, Y., Bourzeix, F., Houmma, I. H., & Hadria, R. (2023). Crop mapping using supervised machine learning and deep learning: A systematic literature review. *International Journal of Remote Sensing*, 44(8), 2717-2753. <https://doi.org/10.1080/01431161.2023.2205984>
- Alami Machichi, M., El Mansouri, L., Imani, Y., Bourja, O., Hadria, R., Lahlou, O., Benmansour, S., Zennayi, Y., & Bourzeix, F. (2022). CerealNet: A hybrid deep learning architecture for cereal crop mapping using Sentinel-2 time-series. *Informatics*, 9(4), 96. <https://doi.org/10.3390/informatics9040096>
- Asam, S., Gessner, U., Almengor González, R., Wenzl, M., Kriese, J., & Kuenzer, C. (2022). Mapping crop types of Germany by combining temporal statistical metrics of Sentinel-1 and Sentinel-2 time series with LPIS data. *Remote Sensing*, 14(2981). <https://doi.org/10.3390/rs14132981>
- Ayushi, & Buttar, P. K. (2024). Satellite imagery analysis for crop type segmentation using U-Net architecture. *Procedia Computer Science*, 235, 3418-3427. <https://doi.org/10.1016/j.procs.2024.04.322>
- Becker-Reshef, I., Barker, B., Whitcraft, A., et al. (2023). Crop type maps for operational global agricultural monitoring. *Scientific Data*, 10, 172. <https://doi.org/10.1038/s41597-023-02047-9>
- Berra, E. F., Fontana, D. C., Yin, F., & Breunig, F. M. (2024). Harmonized Landsat and Sentinel-2 data with Google Earth Engine. *Remote Sensing*, 16(15), 2695. <https://doi.org/10.3390/rs16152695>
- Blickensdörfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasmí, S., & Hostert, P. (2022). Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2, and Landsat 8 data for Germany. *Remote Sensing of Environment*, 269, 112831. <https://doi.org/10.1016/j.rse.2021.112831>
- Brown, M. E. (2014). Food security, food prices and climate variability. Routledge. <https://doi.org/10.4324/9781315773770>
- Choudhury, I., Chakraborty, M., Santra, S. C., & Parihar, J. S. (2011). Methodology to classify rice cultural types based on water regimes using multi-temporal RADARSAT-1 data. *International Journal of Remote Sensing*, 33(13), 4135–4160. <https://doi.org/10.1080/01431161.2011.642018>
- Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., & Justice, C. (2018). The harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, 219, 145–161. <https://doi.org/10.1016/j.rse.2018.09.002>
- Concern Worldwide & Welthungerhilfe. (2024). Global Hunger Index 2024: How gender justice can advance climate resilience and Zero Hunger. Global Hunger Index. <https://www.globalhungerindex.org>
- Copernicus. (n.d.). Copernicus land monitoring service. Retrieved February 5, 2025, from <https://www.copernicus.eu/en/copernicus-services/land>
- Copernicus SentiWiki. (n.d.). S1 Mission. Retrieved May 14, 2025, from <https://sentiwiki.copernicus.eu/web/s1-mission>
- Dadhwal, V. K., Singh, R. P., Dutta, S., & Parihar, J. S. (2002). Remote sensing-based crop inventory: A review of Indian experience. *Tropical Ecology*, 43(1), 107-122.
- Dahiya, N., Singh, G., Gupta, D. K., Kalogeropoulos, K., Detsikas, S. E., Petropoulos, G. P., Singh, S., & Sood, V. (2024). A novel deep learning change detection approach for estimating spatiotemporal crop field variations from Sentinel-2 imagery. *Remote Sensing Applications: Society and Environment*, 33, 101259. <https://doi.org/10.1016/j.rsase.2024.101259>

- Davidson, A. M., Fiset, T., McNairn, H., & Daneshfar, B. (2017). Detailed crop mapping using remote sensing data (Crop Data Layers). In J. Delincé (Ed.), *Handbook on Remote Sensing for Agricultural Statistics* (Chapter 4). *Handbook of the Global Strategy to Improve Agricultural and Rural Statistics (GSARS)*: Rome
- Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaute, T., Savinaud, M., Udrou, C., Valero, S., Bégué, A., Dejoux, J.-F., El Harti, A., Ezzahar, J., Kussul, N., Labbassi, K., ... Koetz, B. (2019). Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. *Remote Sensing of Environment*, 221, 551-568. <https://doi.org/10.1016/j.rse.2018.11.007>
- Desai, G., & Gaikwad, A. (2021). Deep learning techniques for crop classification applied to SAR imagery: A survey. 2021 Asian Conference on Innovation in Technology (ASIANCON), 1-6. <https://doi.org/10.1109/ASIANCON51346.2021.9544707>
- FAO, IFAD, UNICEF, WFP, & WHO. (2024). The state of food security and nutrition in the world 2024 – Financing to end hunger, food insecurity and malnutrition in all its forms. Rome. <https://doi.org/10.4060/cd1254en>
- Fiset, T., Rollin, P., Aly, Z., Campbell, L., Daneshfar, B., Filyer, P., Smith, A., Davidson, A., Shang, J., & Jarvis, I. (2013). AAFC annual crop inventory. <https://doi.org/10.1109/Argo-Geoinformatics.2013.6621920>
- Gajbhiye, K. S., & Mandal, C. (2000). *Agro-Ecological Zones, their Soil Resource and Cropping Systems*. National Bureau of Soil Survey and Land Use Planning, ICAR.
- Gao, F., Jennewein, J., Hively, W. D., Soroka, A., Thieme, A., Bradley, D., Keppler, J., Mirsky, S., & Akumaga, U. (2023). Near real-time detection of winter cover crop termination using harmonized Landsat and Sentinel-2 (HLS) to support ecosystem assessment. *Science of Remote Sensing*, 7, 100073. <https://doi.org/10.1016/j.srs.2022.100073>
- Garnot, V. S. F., Landrieu, L., Giordano, S., & Chehata, N. (2020). Satellite image time series classification with pixel-set encoders and temporal self-attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12325-12334). IEEE.
- Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F., & Brendel, W. (2018). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv*. <https://doi.org/10.48550/arXiv.1811.12231>
- Government of India. (2023). *Agriculture at a glance 2023*. Ministry of Agriculture and Farmers Welfare, Department of Agriculture and Farmers Welfare.
- Government of India, Department of Agriculture and Cooperation, Directorate of Economics and Statistics. (2014). *Operational guidelines for the 12th Five Year Plan (2012-17) Central Sector Scheme of Integrated Scheme of Agriculture Census*, Economics and Statistics. Directorate of Economics and Statistics, Coordination Division, Krishi Bhavan, New Delhi.
- Graesser, J., & Ramankutty, N. (2017). Detection of cropland field parcels from Landsat imagery. *Remote Sensing of Environment*, 201, 165-180. <https://doi.org/10.1016/j.rse.2017.08.027>
- Gumma, M. K., Thenkabail, P. S., Teluguntla, P., Rao, M. N., Mohammed, I. A., & Whitbread, A. M. (2016). Mapping rice fallow cropland areas for short-season grain legumes intensification in South Asia using MODIS 250 m time-series data. *International Journal of Digital Earth*, 9(10), 981-1003. <https://doi.org/10.1080/17538947.2016.1168489>
- Hamidi, M., Homayouni, S., Safari, A., & Hasani, H. (2024). Deep learning-based crop-type mapping using SAR and optical data fusion. *International Journal of Applied Earth Observation and Geoinformation*, 129, 103860. <https://doi.org/10.1016/j.jag.2024.103860>

- Indian Space Research Organisation (ISRO). (2016, December). PSLV-C36 mission overview. Retrieved from https://www.isro.gov.in/media_isro/pdf/Missions/PSLV-C36.pdf
- Ju, Y., & Bohrer, G. (2022). Classification of wetland vegetation based on NDVI time series from the HLS dataset. *Remote Sensing*, 14(9), 2107. <https://doi.org/10.3390/rs14092107>
- Karale, Y. Y., Mohite, J., & Jagyasi, B. (2014). Crop classification based on multi-temporal satellite remote sensing data for agro-advisory services. In T. J. Jackson, J. M. Chen, P. Gong, & S. Liang (Eds.), *Land Surface Remote Sensing II* (Vol. 9260, Article 926004). SPIE. <https://doi.org/10.1117/12.2069278>
- Kawarkhe, M., & Musande, V. (2014). Performance analysis of possibilistic fuzzy clustering and support vector machine in cotton crop classification. In *Proceedings of the 2014 IEEE International Conference on Advanced Communication, Control and Computing Technologies (ICACCCT)* (pp. 114–119). IEEE.
- Koppaka, R., & Moh, T.-S. (2020). Machine learning in Indian crop classification of temporal multi-spectral satellite image. *2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 1–8. <https://doi.org/10.1109/IMCOM48794.2020.9001718>
- Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 778–782. <https://doi.org/10.1109/LGRS.2017.2681128>
- Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 318–327. <https://doi.org/10.1109/TPAMI.2018.2858826>
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166–177. <https://doi.org/10.1016/j.isprsjprs.2019.04.015>
- Mazzia, V., Khaliq, A., & Chiaberge, M. (2020). 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*, 10(1), 238. <https://doi.org/10.3390/app10010238>
- Mahlayeye, M., Darvishzadeh, R., & Nelson, A. (2022). Cropping patterns of annual crops: A remote sensing review. *Remote Sensing*, 14(10), 2404. <https://doi.org/10.3390/rs14102404>
- Mandal, D., Kumar, V., Bhattacharya, A., Rao, Y. S., Siqueira, P., & Bera, S. (2018). Sen4Rice: A processing chain for differentiating early and late transplanted rice using time-series Sentinel-1 SAR data with Google Earth Engine. *IEEE Geoscience and Remote Sensing Letters*, 15(12), 1947–1951. <https://doi.org/10.1109/LGRS.2018.2865816>
- McCormick, R., Thenkabail, P. S., Anece, I., Teluguntla, P., Oliphant, A. J., & Foley, D. (2025). Artificial neural network multi-layer perceptron models to classify California's crops using Harmonized Landsat Sentinel (HLS) data. *Photogrammetric Engineering & Remote Sensing*, 91(2), 91–100. <https://doi.org/10.14358/PERS.24-00072R3>
- NASA Earthdata. (2023, December 7). HLS: Two together are better than two apart. NASA Earthdata. Retrieved from <https://www.earthdata.nasa.gov/news/blog/hls-two-together-better-than-two-apart>
- Nagendram, P. S., Satyanarayana, P., & Teja, P. R. (2023). Mapping paddy cropland in Guntur district using machine learning and Google Earth Engine utilizing images from Sentinel-1 and Sentinel-2. *Engineering, Technology & Applied Science Research*, 13(6), 12427–12432. <https://doi.org/10.48084/etasr.6460>

- National Crop Forecast Centre (NCFC). (n.d.). FASAL: Forecasting Agricultural output using Space, Agrometeorology, and Land-based observations. Government of India. Retrieved March 25, 2025, from <https://www.ncfc.gov.in/fasal.html>
- Ofori-Ampofo, S., Pelletier, C., & Lang, S. (2021). Crop type mapping from optical and radar time series using attention-based deep learning. *Remote Sensing*, 13(22), 4668. <https://doi.org/10.3390/rs13224668>
- Opitz, J. (2024). A closer look at classification evaluation metrics and a critical reflection of common evaluation practice. *Transactions of the Association for Computational Linguistics*, 12, 820–836.
- Orynbaikyzy, A., Gessner, U., Mack, B., & Conrad, C. (2020). Crop type classification using fusion of Sentinel-1 and Sentinel-2 data: Assessing the impact of feature selection, optical data availability, and parcel sizes on the accuracies. *Remote Sensing*, 12(17), 2779. <https://doi.org/10.3390/rs12172779>
- Parreiras, T. C., Bolfe, E. L., Sano, E. S., Victoria, D. C., Sanches, I. D., & Vicente, L. E. (2022). Exploring the Harmonized Landsat Sentinel (HLS) datacube to map an agricultural landscape in the Brazilian savanna. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII–B3–2022, 1113–1118. <https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-1113-2022>
- Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training recurrent neural networks. *arXiv preprint arXiv:1211.5063*. <https://arxiv.org/abs/1211.5063>
- Paul, S., & Kumar, D. N. (2019). Comparison of Landsat-8 and Sentinel-2 data for classification of Rabi crops over Karnataka, India. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (Vol. XLII-3/W6)*. ISPRS-GEOGLAM-ISRS Joint International Workshop on "Earth Observations for Agricultural Monitoring," 18–20 February 2019, New Delhi, India.
- Prashant, P., & Michael, A. (2020). Application of geospatial technology for high-resolution mapping & monitoring of crop patterns in support of crop insurance for the rainfed regions of India. *Proceedings of SPIE - The International Society for Optical Engineering*, 11528, *Remote Sensing for Agriculture, Ecosystems, and Hydrology XXII*; 1152804. <https://doi.org/10.1117/12.2572393>
- Prince, S. J. (2023). *Understanding deep learning*. MIT Press.
- Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). PointNet: Deep learning on point sets for 3D classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 652–660).
- Rahimi-Ajdadi, F., & Khani, M. (2021). Remote sensing-based detection of tea land losses: The case of Lahijan, Iran. *Remote Sensing Applications: Society and Environment*, 22, 100568. <https://doi.org/10.1016/j.rsase.2021.100568>
- Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, 14, 6086. <https://doi.org/10.1038/s41598-024-56706-x>
- Ramachandra, T. V., & Kumar, U. (2004). Spatial decision support system for land use management (Technical Report No. 109). Centre for Ecological Sciences, Indian Institute of Science. https://wgibis.ces.iisc.ac.in/energy/paper/TR109/tr109_mainframe.htm
- Rana, P., & Sathiyamoorthy, V. (2018). Study of the cloud cover variability over the Indian region using geostationary satellite data. *International Archive of Applied Sciences and Technology*, 9(3), 29–34. <https://doi.org/10.15515/iaast.0976-4828.9.3.2934>
- Rao, P., Zhou, W., Bhattarai, N., Srivastava, A. K., Singh, B., Poonia, S., Lobell, D. B., & Jain, M. (2021). Using Sentinel-1, Sentinel-2, and Planet imagery to map crop type of smallholder farms. *Remote Sensing*, 13(10), 1870. <https://doi.org/10.3390/rs13101870>

- Sachs, J. (2015). *The Age of Sustainable Development*. New York Chichester, West Sussex: Columbia University Press. <https://doi.org/10.7312/sach17314>
- Sainte Fare Garnot, V., Landrieu, L., Giordano, S., & Chehata, N. (2020). Satellite image time series classification with pixel-set encoders and temporal self-attention. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 1–10. <https://arxiv.org/abs/2002.02447>
- Senay, G. B., Velpuri, N. M., Bohms, S., Budde, M., Young, C., Rowland, J., & Verdin, J. P. (2020). Drought monitoring and assessment: Remote sensing and modeling approaches for the famine early warning systems network. In V. P. Singh & D. K. Butler (Eds.), *Hydro-Meteorological Hazards, Risks, and Disasters* (pp. 247–276). Elsevier. <https://doi.org/10.1016/B978-0-12-819101-9.00002-9>
- Seydi, S. T., Amani, M., & Ghorbanian, A. (2022). A dual attention convolutional neural network for crop classification using time-series Sentinel-2 imagery. *Remote Sensing*, 14(3), 498. <https://doi.org/10.3390/rs14030498>
- Sher, M., Minallah, N., Frnda, J., & Khan, W. (2024). Elevating crop classification performance through CNN-GRU feature fusion. *IEEE Access*, 12, Article 3467193. <https://doi.org/10.1109/ACCESS.2024.3467193>
- Sheykhoumousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 6308–6325. <https://doi.org/10.1109/JSTARS.2020.3026724>
- Singh, G., Singh, S., Sethi, G., & Sood, V. (2022). Deep learning in the mapping of agricultural land use using Sentinel-2 satellite data. *Geographies*, 2(4), 691–700. <https://doi.org/10.3390/geographies2040042>
- Singh, V., Jose, J., & Kalubarme, M. H. (2013). Village-level crop inventory using remote sensing and geo-informatics in Gujarat State, India. *Asian Journal of Geo-informatics*, 13(2), 31–38.
- Song, X.-P., Huang, W., Hansen, M. C., & Potapov, P. (2021). An evaluation of Landsat, Sentinel-2, Sentinel-1 and MODIS data for crop type mapping. *Science of Remote Sensing*, 3, 100018. <https://doi.org/10.1016/j.srs.2021.100018>
- Sivaraj, N., Pandravada, S. R., Kamala, V., Sunil, N., Rameash, K., Abraham, B., Elangovan, M., & Chakrabarty, S. K. (2013). Indian crop diversity. In M. Elangovan & J. V. Patil (Eds.), *Managing intellectual property under PVP and PGR* (pp. 83–88). Directorate of Sorghum Research (DSR).
- Teluguntla, P., Ryu, D., George, B., Walker, J. P., & Malano, H. M. (2015). Mapping flooded rice paddies using time series of MODIS imagery in the Krishna River Basin, India. *Remote Sensing*, 7(5), 5010–5036. <https://doi.org/10.3390/rs70505010>
- United Nations. (2024). *World population prospects 2024: Summary of results* (UN DESA/POP/2024/TR/NO. 9). New York: United Nations.
- United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS). (2024). *Cropland Data Layer*. USDA NASS Marketing and Information Services Office, Washington, D.C. Retrieved from <https://croplanderos.scinet.usda.gov/>
- UNOOSA and FAO. (2025). *Leveraging space technology for agricultural development and food security*. Vienna. <https://doi.org/10.4060/cd3989en>
- U.S. Geological Survey. (2015). *Landsat—Earth observation satellites* (ver. 1.4, August 2022). U.S. Geological Survey Fact Sheet 2015–3081, 4 p. <https://doi.org/10.3133/fs20153081>

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)* (pp. 6000–6010). Curran Associates Inc., Red Hook, NY, USA.
- Victor, B., Nibali, A., & He, Z. (2025). A systematic review of the use of deep learning in satellite imagery for agriculture. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18, 2297–2316. <https://doi.org/10.1109/JSTARS.2024.3501216>
- Vijayan, D. V., Madane, D. A., & Haldar, D. (2023). Monsoon paddy crop discrimination using machine learning algorithms to multi-temporal Sentinel-1A (C-band) data in Alathur block of Palakkad district of Kerala state, India. *Paddy and Water Environment*, 21, 365–375. <https://doi.org/10.1007/s10333-023-00934-w>
- Wang, S., Di Tommaso, S., Faulkner, J., Friedel, T., Kennepohl, A., Strey, R., & Lobell, D. B. (2020). Mapping crop types in Southeast India with smartphone crowdsourcing and deep learning. *Remote Sensing*, 12(18), 2957. <https://doi.org/10.3390/rs12182957>
- Yan, Y., & Ryu, Y. (2019). Google Street View and deep learning: A new ground truthing approach for crop mapping. *arXiv*. <https://arxiv.org/abs/1912.05024>
- You, L., & Sun, Z. (2022). Mapping global cropping system: Challenges, opportunities, and future perspectives. *Crop and Environment*, 1(1), 68–73. <https://doi.org/10.1016/j.crope.2022.03.006>
- Weilandt, F., Behling, R., Goncalves, R., Madadi, A., Richter, L., Sanona, T., Spengler, D., & Welsch, J. (2023). Early crop classification via multi-modal satellite data fusion and temporal attention. *Remote Sensing*, 15(3), 799. <https://doi.org/10.3390/rs15030799>
- Zhao, W., Deledalle, C.-A., Denis, L., Maître, H., Nicolas, J.-M., & Tupin, F. (2018). Ratio-based multitemporal SAR images denoising: RABASAR. *IEEE Transactions on Geoscience and Remote Sensing*, 56(6), 3442–3454. <https://doi.org/10.1109/TGRS.2018.2809602>
- Zoungrana, L. E., Barbouchi, M., Toukabri, W., Ould Babasy, M., Ben Khadra, N., Annabi, M., & Bahri, H. (2023). Sentinel SAR-optical fusion for improving in-season wheat crop mapping at a large scale using machine learning and the Google Earth Engine platform. *Applied Geomatics*, 16, 147–160. <https://doi.org/10.1007/s12518-023-00545-4>

Appendix

Appendix A: Survey planning and field survey

Survey planning

The survey was meticulously planned using geographical and crop type knowledge acquired from talking to local farmers. Planning was dynamic in nature, but we stuck to the designed routes. Local contacts and hosting families did contribute to the planning of the routes. Over a period of 52 days, extensive fieldwork was conducted around more than 40 villages, covering over 2,500 kilometers by driving and 300+ kilometers on foot. This survey spanned across five districts, involving diverse terrains, cropping pattern, practices and different agricultural systems. During these surveys we started with the roadside transect sampling method but later figured out two other methods, all of which are discussed below.

Type of sampling method followed

1. Roadside Transect sampling: This is the easiest one. In this method we drive along the rural road and captured the crop type field polygon on either side of the road Image A of the Figure 14 shows the red line indicating the road and marked polygons on the either side of the road. This method is core of our sampling technique.
2. L-shaped Transect sampling: It is hybrid transect method, where sampling follows a linear path (e.g., along a road) and then turns into the village roads as shown in image B of Figure 14. This is useful to capture variations in crop type influenced by proximity to roads.
3. U-shaped Transect Survey (name given by author): It is hybrid of systematic and exploratory sampling. It enabled us to collect a greater variety of crop data and penetrate deeper into rural areas. Unlike traditional linear transects, this method involved a “U-shaped” or sometimes “O-shaped” walking pattern. Surveyor would begin at a fixed starting point and walk sequentially through 3 to 4 adjacent fields in a straight line. Upon reaching the end of this segment, surveyor would then turn perpendicularly and continue walking through another 4 to 5 fields. This pattern would repeat, until the surveyor returned to the starting point. This survey was discovered and followed in the latter period of surveying. Image C of the Figure 14 shows red line as the path followed by surveyor during the collection of data.

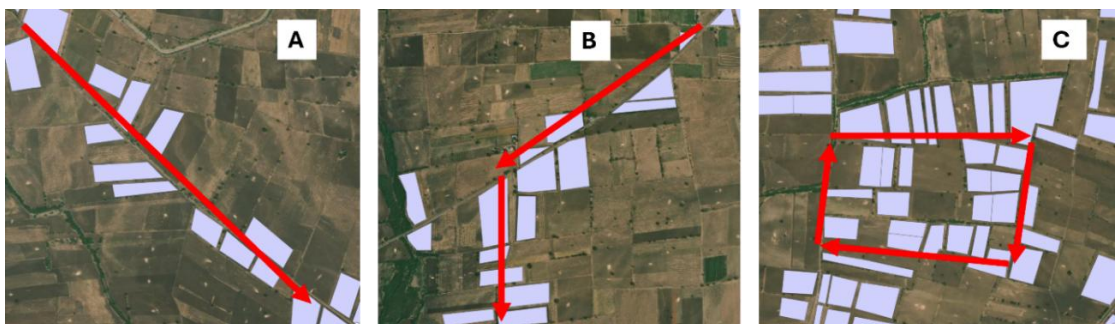


Figure 14. Figure describes three survey methods. Image (A) shows Roadside Transect sampling, image (B) shows L-shaped Transect sampling and image (C) shows U-shaped Transect sampling.

Field survey

Figure 15 shows the sequence illustrating crop surveying using Esri's Field Maps application. The process begins with the user selecting a specific crop type (image A) from a list — in this case, **Wheat** under the category of agriculture crops. Once selected, the application opens a GPS-enabled map view that allows the user to collect data. The user then adds points to digitize the field boundary (image B).

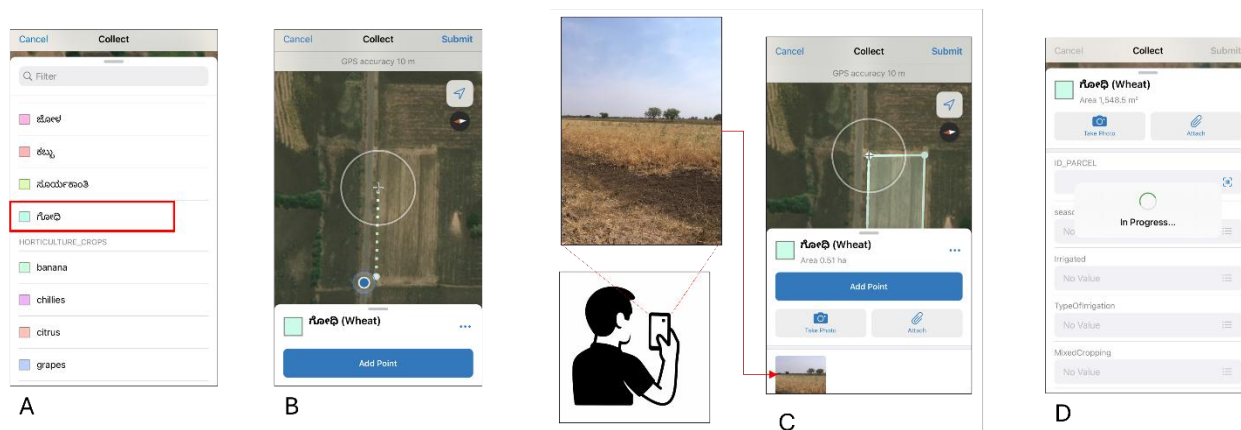


Figure 15. The figure illustrates key steps in crop type data collection at the field level: crop selection (image A), field marking on the base map using the polygon drawing tool (image B), capturing an oriented image post polygon completion (image C), and submission process (image D).

Adding polygon can be done either by physically walking along the field or by referring to visible landmarks such as trees, field boundaries, or nearby roads that appear clearly on the basemap and match the real-world surroundings. Using visual correlation is an important part in this survey where a structure can be validated on high resolution base maps. Figure 16 shows a real-world scenario where the tree captured in the photograph is clearly visible on satellite imagery, confirming the spatial accuracy of the field observation.

After drawing the field geometry and adding relevant attribute information, the user finalizes the entry by tapping the Submit button (Submit button exists in the right corner of app as shown in image C). Image D shows the “In Progress” indicating that data is being sent to cloud storage . The GPS accuracy is shown as 10 meters, which is a threshold set by the user to ensure spatial precision during data collection.

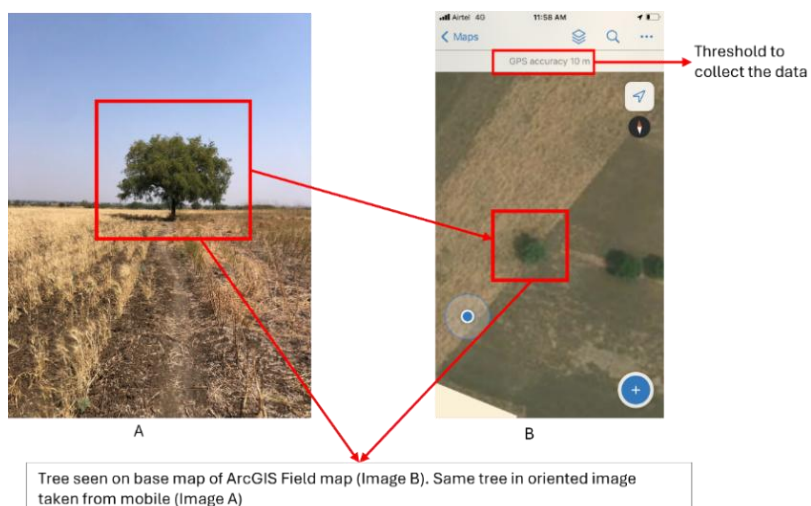


Figure 16. Figure shows a tree bordering the wheat crop field. Image (A) shows the tree as a ground reference with respect to same reference on the Esri's Base Map (image B).

The collected data is uploaded in real time to the ArcGIS datastore. Quality analysis involving checking the geometry and attributes is done using the published feature service using ArcGIS pro software. Daily or weekly data backup can also be taken based on the requirement. We maintained 10 days backup in the form of shapefiles.

For the real time monitoring and daily planning, a mobile compatible dashboard was designed using ArcGIS desktop. This dashboard helped with keeping tabs on the number of field visits. Figure 17 shows the mobile compatible dashboard showing the number of horticultural and agricultural crop types captured. We also introduced pie charts for both agricultural and horticultural crops. These pie charts helped use to maintain the number ratio between various crop types.

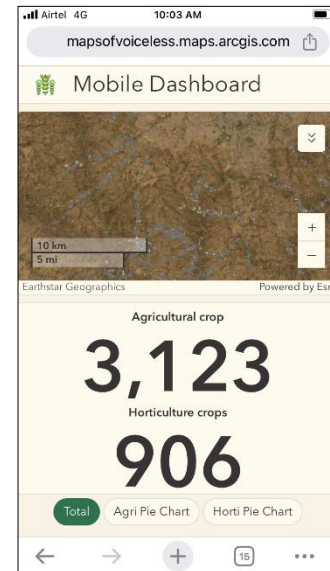


Figure 17 : Mobile compatible dashboard.

Appendix B: Observed crop mixing

Observed crop mixing

Sorghum is a versatile and hardy crop grown both as animal fodder and as a food grain. In the study area (Regions A and B), it is mostly cultivated during the rabi season. While there were fields where only sorghum was grown, it was commonly found as a mixed crop - especially with chickpea, and occasionally with wheat. The mixing pattern was often systematic, typically consisting of a single row of sorghum followed by five to seven rows of chickpea, as shown in image A of Figure 18. Sorghum was also observed planted along one edge or around all edges of the field, as illustrated in image B.

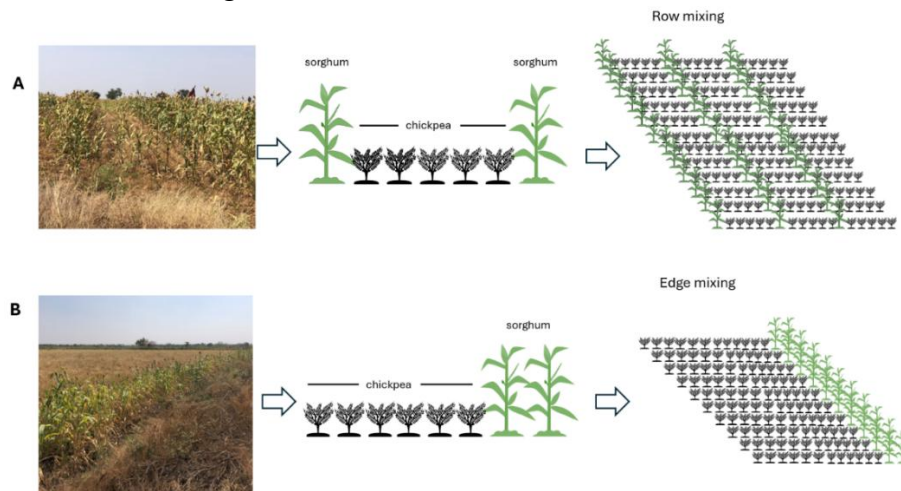


Figure 18. Mixing patterns of sorghum with chickpea. Image A shows cropping pattern containing row mixing. Image B shows edge mixing pattern.

In some fields, only a few strands of sorghum were present with chickpea, as shown in image A of Figure 19. In image A, harvested chickpea is visible, with a few unharvested strands of sorghum still standing.

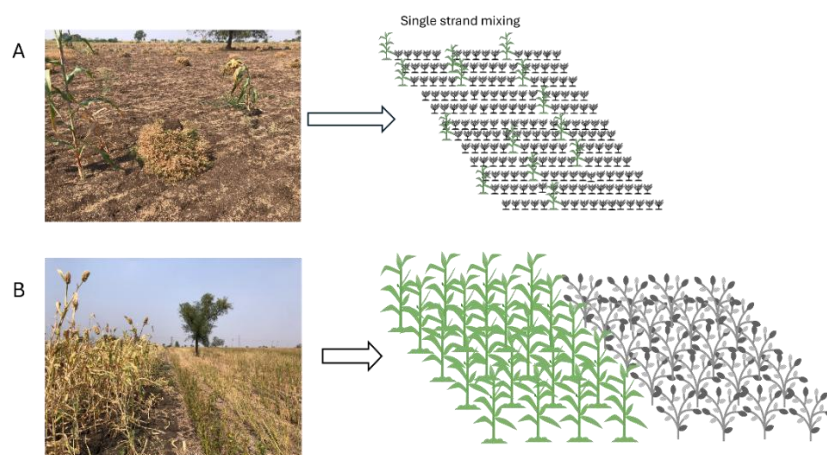


Figure 19. More sorghum crop mixing patterns. Image A shows representation of single strand mixing pattern. Image B shows plot level mixing pattern.

We also observed plot-wise mixing of crops, where a single field contained two or three different crops, each occupying a substantial area but without any physical boundaries separating them as

shown in image B of Figure 19. Sampling such fields as polygons is particularly challenging due to the risk of spectral mixing, which can result in mixed pixels. To minimize this issue, polygons were drawn slightly away from the observed crop boundaries on the base map, ensuring that pure pixels were captured as much as possible.

Another pattern we observed was related to waterlogging. In areas where the original crop was destroyed due to waterlogging, farmers had replaced the affected sections with either wheat or chickpea. These replacements were usually in small patches and did not cover a substantial portion of the field. In cases where such patchy mixing was significant, the sample was discarded. Figure 20 illustrates how this type of mixing occurs.

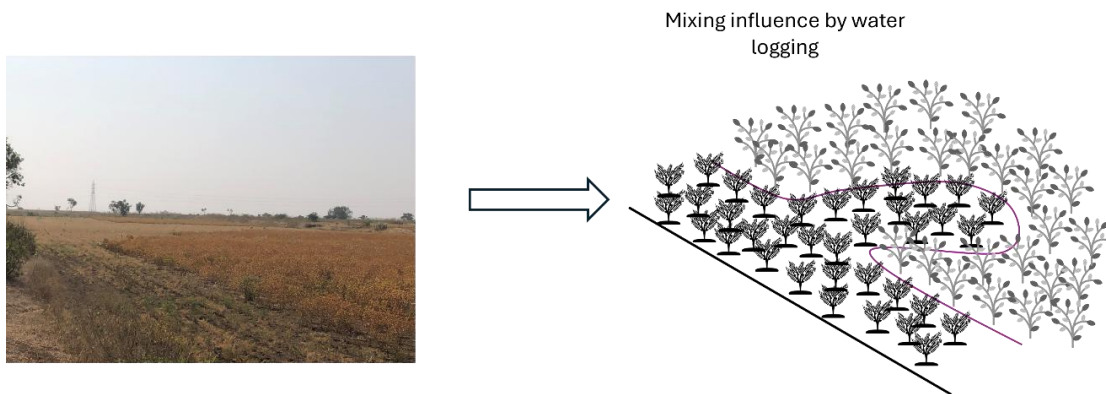


Figure 20. Crop mixing as replacement for destroyed crops.

In Region B, an interesting crop mixing pattern was observed where sugarcane was grown amidst mango plantations. This was a common practice in mango-growing areas, and therefore, polygons containing such mixed crops were included in the Region B dataset. Figure 21 illustrates two types of sugarcane-mango mixing: Top image shows mixing along the field boundary, while bottom image depicts sugarcane planted between mango trees.

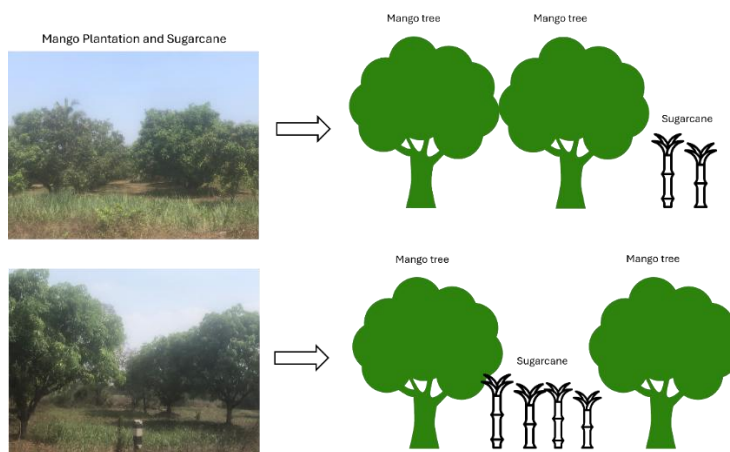


Figure 21. Mixing of sugarcane in mango plantations.

Appendix C: Cloud cover statistic and filtering

Cloud cover statistic

The analysis of cloud cover (as in Table 11) for three satellite datasets (HLS L30, HLS S30, and Sentinel-2) revealed significant seasonal variations in cloud cover. The January-March 2025 period showed substantially lower cloud cover across all optical sensors compared to September-December 2024. Notably, Sentinel-2 showed the most dramatic improvement, with average cloud cover dropping from 46.3% to just 2.8% between the two periods. HLS L30 also showed significant improvement from 30.7% to 10.7%, while HLS S30 showed more modest improvement from 36.1% to 31.8%. These patterns are consistent with typical seasonal weather patterns in the region, with clearer conditions prevailing in the early months of the year.

Table 11. Cloud cover statistics for Sentinel-2, HLSS30 and HL30 datasets.

Time Period	Sentinel-2	HLSS30	HL30
September – December 2024	46.3%	36.1%	30.7%
January – March 2025	2.8%	31.8%	10.7%

Cloud cover filtering

Cloud cover percentage in the satellite tiles is one of the key parameter that will define how noisy the data will be. More noisy the data, less accurate the classification. Following data experiment were conducted to determine the cloud cover percentage threshold below which the satellite data provides optimal classification accuracy.

For HL30 dataset, three dataset were extracted with cloud cover <20 %, cloud cover <50% and cloud cover <100% (no cloud filter). PSETAE model was run on these three different dataset by keeping the same hyperparameters. Results are as given in Table 12. Though there was no much difference in accuracy with respect to these three results in terms of accuracy. There is improvement in Mean IoU when cloud cover goes on decreasing. Based on accuracy and mean IoU results we concluded to keep cloud cover filter less than 20 for data extraction. This will also decrease the processing time needed as amount of data will be less compared to cloud cover < 100%. Similarly for HLS-S30 we decide apply Cloud cover < 20%.

Table 12. Results of cloud cover experiment for HL30.

Metrics	Cloud cover < 100 %	Cloud cover < 50 %	Cloud cover < 20 %
Accuracy in %	87.12	87.50	87.50
Mean IoU	0.657	0.674	0.683

For Sentinel-2 data, extraction of data was carried out keeping cloud cover <100 % (i.e. no cloud filter), cloud cover <20% and cloud cover <10% . PSETAE model was run on these dataset

keeping the same hyperparameter. Results are as given in Table 13. There was minor change in the accuracy and Mean IoU. Interesting part of this experiment is that cloud cover < 100% gave better results than cloud cover<20% and cloud cover<10%. This could be attributed to enormous amount of the data that gets added if all the tiles are taken. As there was less difference between the results amount to just 1% and to maintain similarity in the experiments we decide to go with cloud cover <20% for Sentinel-2 data as well.

Table 13. Results of cloud cover experiments for Sentinel-2.

<i>Metrics</i>	<i>Cloud cover < 100 %</i>	<i>Cloud cover < 20 %</i>	<i>Cloud cover < 10 %</i>
<i>Accuracy in %</i>	96.08	95.29	94.96
<i>Mean IoU</i>	0.866	0.853	0.847

Appendix D: PSETAE equations

Pixel Set Encoder

PSE can be summarized by following equations.

$$s = \text{Sample}(S, N)$$

$$\hat{e}_s^{(t)} = \text{MLP}_1(x_s^{(t)})$$

$$e^{(t)} = \text{MLP}_2([\text{pooling}(\{\hat{e}_s^{(t)}\}) + G \text{ info}])$$

Temporal Attention Encoder

TAE can be summarized by following equations

$$k_h^{(t)}, q_h^{(t)} = \text{FC}_1^{(h)}(e^{(t)} + p^{(t)})$$

$$\hat{q}_h = \text{FC}_2^{(h)}(\text{mean}(\{q_h^{(t)}\}_{t=1}^T))$$

$$a_h = \text{softmax}(\frac{1}{\sqrt{d_k}} [\hat{q}_h * k_h^{(t)}]_{t=1}^T)$$

$$o_h = \sum_{t=1}^T a_h[t](e^{(t)} + p^{(t)})$$

$$\hat{o} = \text{MLP}_3([o_1, o_2, \dots, o_H])$$

$$y = \text{MLP}_4(\hat{o})$$

$$y_{\text{class}} = \text{argmax}(y)$$

s is 1 random pixel taken from S number of pixels from total number of N pixels in the parcel.

x can be expanded as  where b_1, b_2 etc. are band values. t is observation time of time-series

MPL is multi-layer perceptron. $e^{(t)}$ is the parcel's spatial-spectral embedding at time t . $G \text{ info}$ is geometric feature information of the parcel that will be concatenated. $+$ represents the concatenation

$p^{(t)}$ is positional encoding given by

$$[p^{(t)}]_{i=1}^{d_e} = \sin\left(\text{day}(t) \setminus \tau^{\frac{2i}{d_e}} + \frac{\pi}{2} \text{mod}(i, 2)\right)$$

$k_h^{(t)}, q_h^{(t)}$ are key-query pairs from the spatial embeddings. $\text{FC}_1^{(h)}$ is the fully connected layer generating the key query pairs for each head h .

$\hat{q}_h$ is master query. $\text{FC}_2^{(h)}$ is fully connected layer that yield master query. d_k is the shared dimensions of keys and queries.

o_h is the attention output summarized for $t \in [1, \dots, T]$

$\hat{o}$ is the resulting embedding for $h \in [1, \dots, H]$. Where h are attention heads. MPL3 is multi-layer perceptron.

y is the processed class logits from MLP decoder. For more clarity, I have introduced y class which give the class number after taking raw logits. Softmax can be applied to get class probabilities.

PSETAE Schematic view by Garnot et al. (2020)

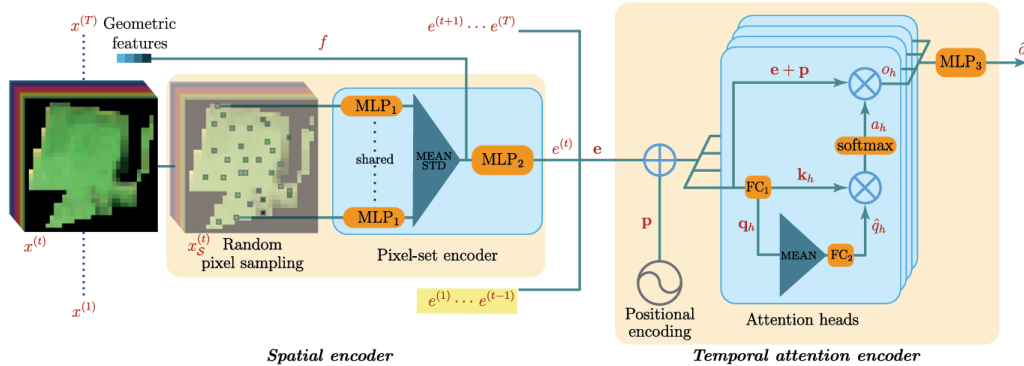


Figure 22. PSETAE architecture. Source: Garnot et al. (2020).