Crop Mapping Using Deep Learning and Multi-Source Satellite Remote Sensing

Niantang Liu

SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

SCHOOL OF GEOGRAPHICAL AND EARTH SCIENCES

COLLEGE OF SCIENCE AND ENGINEERING

UNIVERSITY OF GLASGOW

JANUARY 2024

Abstract

Crop mapping is the prerequisite process for supporting decision-making and providing accurate and timely crop inventories for estimating crop production and monitoring dynamic crop growth at various scales. However, in-situ crop mapping often proves to be expensive and labour-intensive. Satellite remote sensing offers a more cost-effective alternative that delivers time-series data that can repeatedly capture the dynamics of crop growth at large scales and at regularly revisited intervals. While most existing crop-type products are generated using remote sensing data and machine learning approaches, the accuracy of predictions can be low given that misclassifications persist due to phenological similarities between different crops and the complexities of farming systems in real-life scenarios. Deep neural networks demonstrate great potential in capturing seasonal patterns and sequential relationships in time series data in the context of their end-to-end feature learning manner. This thesis presented a comprehensive exploration of advanced deep learning methodologies for large-scale agricultural crop mapping using multi-temporal and multi-source remote sensing data. Focusing on Bei'an County in Northeast China, the research developed and evaluated innovative frameworks to produce accurate crop-specific map products, addressing challenges such as optimal satellite-based input feature selection, imbalanced crop type distribution, model transferability, and model learning visualisation. This research has effectively addressed these challenges in complex agricultural environments by introducing advanced deep learning architectures that utilise multi-stream models and multi-source data fusion. The classification frameworks developed through this thesis have shown improved performance in accurately mapping crops, particularly in terms of evaluating model generalisability for inference of unseen area, model spatial and interannual transferability across different test sites, and model interpretability for unveiling the model decision process that contributes to a deeper understanding of model learning behaviours for temporal growth patterns of crops. The findings highlight the importance of temporal dynamics, the integration of various data sources, and the effectiveness of ensemble learning in enhancing the accuracy and reliability of crop classification. A deep learning framework using radar-based features was developed, achieving F1 scores for maize (87%), soybean (86%), and other crops (85%) on an imbalanced crop dataset. This approach was extended by integrating Sentinel-1 and Sentinel-2 data, resulting in an overall accuracy of 91.7%, with F1 scores of 93.7%, 92.2%, and 90.9% for maize, soybean, and wheat, respectively. Furthermore, the spatiotemporal transferability of pre-trained models was systematically evaluated across two test sites, resulting in overall accuracies of 96.2% and

90.7%, mean F1 scores of 92.7% and 88.6%, and mean IoUs of 86.9% and 79.7% for site A and site B, respectively.

Acknowledgements

Embarking on this academic journey that led me to study at the University of Glasgow for my MSc and, subsequently, my PhD has been a long journey in my life, but this sets the tone for the beginning of my career. As Steve Jobs said, ‘Stay Hungry, Stay Foolish,’ my exploration of creativity continues. To complete this tough journey, which I often playfully characterise as a mental purgatory, has been made possible by the sustained support of numerous individuals and organizations, allowing me to grow through failure, learn from it, and maintain patience.

First and foremost, I would like to send my heartful thanks and appreciation to my primary supervisor, Dr Brian Barrett. He guided me towards each milestone with patience and encouragement during my PhD life. I am particularly grateful for his understanding whenever I made mistakes, and for the energy I gained from our in-person meetings, where he wisely addressed my concerns from various aspects. Brian’s support in presenting my research at the conferences held by the Remote Sensing and Photogrammetry Society (RSPSoc) for consecutive years, 2022 and 2023, boosted my confidence to deliver oral presentations. My secondary supervisors, Dr Qunshan Zhao and Professor Richard Williams, also deserve my sincere appreciation for their professional advice, feedback, and technical support, which significantly improved my research and writing skills. Their collective efforts have shaped me into a competent researcher, capable of contributing to academic journals. The pandemic presented unforeseen challenges to my research progress. Brian’s initiative to regulate weekly and monthly meetings for our group provided much-needed supervision and guidance, which helped me regain my focus. The persistent efforts and alternative research schemes suggested by every supervisor during our meetings were meaningful to me and essential in directing me back on the right track. I would also like to express my gratitude to Dr Jiren Xu, serving as my internal examiner, and Professor Kevin Tansey as my external examiner during the Viva. Their suggestions and support as to my thesis and future research directions have been informative.

I owe my deepest gratitude to my family, friends, and girlfriend, whose support has been my spiritual strength. They have been there for me during times of uncertainty and exhaustion, motivating me with every phone call. My father, who sometimes acted as my external supervisor, often engaged with me in discussions beyond the academic aspects. Starting with my recent research dilemma, he gradually brought me into philosophical realms and extended

relevant topics. His profound influence has reshaped my perspective, leading me to rethink the workaround in a broad context.

Through doing extensive research and experiments, I have eventually come to understand and appreciate certain philosophical truths. For instance, some things that appear incredibly complex are realistically achievable, while others that seem straightforward are, in reality, complicated. To clarify, when I revisited literature that I found difficult to understand from my early PhD years, I realised that these were a series of simple steps but intricately linked together in hierarchical manners. Ensuring that each of these 'simple' steps is completed correctly and can be justifiably connected, however, is far from easy.

Lastly, I would like to thank the Chinese Academy of Agricultural Sciences (CAAS) for providing the important ground truth dataset that facilitated much of my research.

Niantang Liu

January 2024, Glasgow

Declaration

I declare that I am the sole author of the work contained within this thesis, except where explicit reference is made to the contribution of others, and that it is of my own composition. No part of this work has been submitted for any other degree at the University of Glasgow or any other institution.

Niantang Liu

January 2024, Glasgow

Abbreviations

AtLSTM	Attention-Based Long Short-Term Memory
CDL	Cropland Data Layer
CNNs	Convolutional Neural Networks
ConvRNN	Convolutional Recurrent Neural Networks
ConvSTAR	Convolutional STAckable Recurrent Cell
DpRVI	Dual-pol Radar Vegetation Index
FCNs	Fully Convolutional Networks
GLCM	Grey Level Co-occurrence Matrix
InSAR	Interferometric Synthetic Aperture Radar
LSTM	Long Short-Term Memory
1D-CNNs	One-dimensional Convolutional Neural Networks
Conv1D	One-dimensional Convolution
PolSAR	Polarimetric Synthetic Aperture Radar
RF	Random Forest
RNNs	Recurrent Neural Networks
SAR	Synthetic Aperture Radar
STAR	STAckable Recurrent Cell
2D-CNNs	Two-dimensional Convolutional Neural Networks
3D-CNNs	Three-dimensional Convolutional Neural Networks

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background

In alignment with the 2030 Agenda for Sustainable Development, the United Nations (UN) established 17 Sustainable Development Goals (SDGs), among which "zero hunger" stands prominent. This specific goal aims to tackle the challenges posed by the rapid growth of the global population, as projected to reach 8.54 billion by 2030, and the consequent surge in global food demand. Concurrently, the Food and Agriculture Organization (FAO) has emphasized the impacts of these challenges on the long-term sustainability of agricultural systems (UN General Assembly, 2015; Desa, 2019; FAO, 2018). Crop production and food security remain challenges in our society, necessitating an efficient and sustainable agricultural framework to balance the rising demand for feedstock crops while also contending with the impacts of climate change (Ray et al., 2013). Maize and soybean, as dominant commodity crops in both national and international markets, are the requisite sources in the global food supply chain (Wu et al., 2021). Specifically, the United States and China stand out as the primary maize producers, together accounting for over half of worldwide maize production (You et al., 2023). Therefore, obtaining accurate and timely information on sowing and harvesting areas, as well as the production and yield of crops such as maize and soybean, are essential for decision-making, crop growth monitoring, yield prediction, acreage estimation, food security, and facilitating international trade (Carfagna and Gallego, 2005; Tilman et al., 2011; Iizumi and Ramankutty, 2015; Wang et al., 2019b).

1.2 Advances in Satellite-Based Crop Monitoring

Agricultural information has traditionally been collected through census and field surveys (Song et al., 2017). Field surveys provide data that can be characterised as indicators for diverse agricultural facets, such as cropland size, cultivated regions, land ownership, fertilizer application, labour, and irrigation practices, while censuses are typically conducted once a decade, making them more suited for tracking slow-evolving trends of agriculture (FAO, 2015). In addition, a challenge with using census data on a global scale can lead to inconsistencies, which can arise from varied definitions of census metrics, changing political boundaries, and diverse reporting protocols across countries and census periods (Portmann et al., 2010). While fieldwork can yield high-quality data, its labour-intensive nature poses difficulties in practice (Zhang et al., 2021). Satellite-based remote sensing has increasingly been applied in

operational agricultural surveys, primarily due to its enhanced spatial coverage and consistent revisit capabilities. This approach delivers up-to-date and spatially contiguous data, facilitating the monitoring of crop dynamics, sown areas, and overall crop yield at both regional and global scales since the early 1970s (Lobell 2013; Fritz et al. 2019). This data is ever more openly accessible e.g. from coarse spatial resolution 500-m Moderate Resolution Imaging Spectroradiometer (MODIS) data (Massey et al. 2017; Chen et al., 2018), to the medium 30-m resolution Landsat data (Cai et al., 2018; Wang et al., 2019c; Zhong et al., 2014; Dong et al., 2016; Johnson, 2019; Oliphant et al., 2019; Wen et al., 2022) and the finer 10-m resolution of Sentinel-2 and Sentinel-1 data (Belgiu and Csillik, 2018; You et al., 2021; Maponya et al., 2020; Gallo et al., 2023; Wei et al., 2021; Ni et al., 2021). Additionally, the synergistic use of multi-source datasets has showcased the potential of data fusion, offering opportunities for advanced research in crop classification across multiple scales (Adrian et al., 2021; Blickensdörfer et al., 2022; Wang et al., 2022). Nonetheless, utilising satellite data to map specific crop types throughout the entire growth cycle or across different growth stages remains challenging, given the complexities of cropping systems, which include diverse crop types with similar spectral features, cropping patterns, cropland sizes and management practices.

Machine Learning techniques and their advanced derivative, Deep Learning (LeCun et al., 2015), have emerged to enhance more sophisticated and nuanced data interpretations further. Recently, the Geospatial Artificial Intelligence (GeoAI) field in the Earth observation (EO) community has gained notable progress, particularly in large-scale prediction tasks such as satellite imagery classification and global climate modelling (VoPham et al., 2018; Shi et al., 2023). Machine learning algorithms have been globally employed on national scales to automate the extraction of meaningful information from diverse geospatial data sources (Jin et al., 2019; Wang et al., 2020; Pott et al., 2021). In contrast, deep learning approaches use their inherent capabilities to alleviate the necessity for manually engineered data to collect spatial, temporal, or spectral features, which is the procedure commonly required in machine learning approaches such as support vector machines and multilayer perception. Deep learning models, which utilise complex neural network architectures, can efficiently extract complicated and non-linear relationships within the high-dimensional data in remotely sensed images, provided the prerequisites are met in terms of the availability of a large amount of labelled training data and sufficient computational resources.

1.3 Deep Learning in Remote Sensing

Convolutional Neural Networks (CNNs) have been a predominant deep learning architecture for remote sensing applications. Their computational efficiency and robustness to learning features position them as superior models for recognizing structures and patterns in multi-dimensional data. CNNs inherently adopt a hierarchical approach to feature representation within image datasets, progressing from individual pixels forming edges, which combine to create motifs, further assembling into parts, leading to objects, and culminating in the representation of entire scenes. As such, this multi-layered structure of CNNs facilitates the extraction of intricate feature representations from multi-source image datasets (Zhang et al., 2016). The increasing adoption of CNNs across diverse remote sensing challenges and various satellite data types (e.g. Ma et al., 2019; Kattenborn et al., 2021) has demonstrated their abilities in applications including land cover classification (e.g. Zhang et al., 2019; Shendryk et al. 2019; Li et al., 2022; Mazzia et al., 2019), object detection (e.g. Chen and Wang, 2014; Zhao et al., 2019), and semantic segmentation (e.g. Wei et al., 2019; Adrian et al., 2021).

CNNs have evolved into several variants for enhanced learning efficiency. For instance, one-dimensional CNNs (1D-CNNs) focus on the extraction of temporal or spectral features based on single-pixel, without considering spatial inter-pixel relationships and context information (Kiranyaz et al., 2021). In contrast, typical 2D-CNNs segment large input images into smaller patches with each label corresponding to the centre pixel of each patch (Sharma et al., 2017). However, this patch-based approach could lead to uncertainties in the classification results in terms of boundary artefacts between patches, potentially leading to over-smoothness of the object boundaries, which in turn compromises the clarity and precision of the predicted objects (Zhang et al., 2018). While overlapping patches can mitigate some of these issues, they introduce redundant information and cause computational burdens. On the other hand, three-dimensional CNNs (3D-CNNs) offer a more comprehensive feature learning, considering both spatial and temporal dimensions (Ji et al., 2018; Mäyrä et al., 2021; Gallo et al., 2023). Fully Convolutional Networks (FCNs) represent another deep learning paradigm that is particularly adept at handling remote sensing imagery, as evidenced by their successful image semantic segmentation in land cover classification and crop mapping (Mohammadimanesh et al., 2019; Li et al., 2021). Unlike CNNs with small patches, FCNs consider wider contexts of input images and retain the full dimensionality of input and output images. However, their

effectiveness hinges on the availability of high-quality labelled training data for each pixel of the imagery.

Recurrent Neural Networks (RNNs) are a subset of neural networks suitable for recognising recurring patterns, making them particularly advantageous for multi-temporal remote sensing analysis. Compared to conventional neural networks, RNNs are characterised by their unique structure that consumes recursive information, enabling them to efficiently capture sequential correlations by linking successive input variables (Werbos, 1990). Tailored for the analysis of sequential data, RNNs have proven their capabilities across a range of remote sensing applications, especially for crop classification (Zhong et al., 2019; Xu et al., 2020; Rußwurm and Körner, 2020; Turkoglu et al., 2021). Their intrinsic mechanism to process data with sequential dependencies makes RNN-based deep learning models optimal for capturing temporal relationships in observations and modelling change dynamics of objects within time series remote sensing images (Mou et al., 2018). This is also particularly useful in agricultural domains in terms of their ability to understand and model the temporal behaviour of crops, including their growth stages, phenological changes, and responses to environmental factors.

This dynamism of crop objects, characterized by their seasonal growth patterns, phenological developments, and varying spectral signatures over time, presents a complex challenge that temporal models are designed to handle. By analysing the temporal sequences of remote sensing data, some deep learning models including RNNs are adept at identifying and predicting complex crop dynamics, thereby improving agricultural cropland mapping. By leveraging their capability for sequential data processing, these models can effectively learn and extract meaningful patterns within the temporal progression of satellite imagery, which makes them better understand the implications of previous growth stages on current crop conditions. Among the various RNN configurations, the Long Short-Term Memory (LSTM) model is one of the variants that incorporate gating mechanisms to reveal long-term dependencies in time series data sequences (Hochreiter and Schmidhuber, 1997). Therefore, the multi-layered design of the LSTM model is also well-suited for handling multi-temporal satellite observations (Boulila et al., 2021; Chen et al., 2022).

While the aforementioned deep learning architectures have shown remarkable success in various remote sensing applications, several challenges persist. These include handling the vast and growing volume of satellite data with advanced model architectures and improving the

accuracy and reliability of predictions across diverse cropping patterns. The next section, the problem statement, introduces these aspects in greater detail, by exploring the current limitations of crop classification tasks and identifying key areas where further research and development are essential to advance this domain.

1.4 Problem Statement

Accurate and detailed agricultural land cover maps are essential for developing sustainable agriculture plans on a global scale by serving as critical inputs to simulation models that can evaluate environmental and socioeconomic changes (Blickensdörfer et al., 2022; Jin et al., 2018). These maps provide comprehensive spatial cropland distributions over multiple years, facilitating the analysis of cropland management and reflecting land use intensity. In Northeast China, a key agricultural region for the country, the local cropping systems are subject to annual changes due to practices like crop rotation and the soybean rejuvenation plan, which aim to foster sustainable farming, alleviate trade pressures on specific crops, and optimise land allocation strategies (Yang et al., 2019; Guo et al., 2021). Despite these initiatives, annual crop maps on a large scale are rarely available for this region, which limits the ability to quantitatively assess changes in local farming systems and the understanding of cropland dynamics.

In many agricultural sectors within local Chinese governments, labour-intensive manual labelling tasks are undertaken to ensure accurate annual reporting (Ji et al., 2018). Given that annual crop inventory data is used for both agricultural applications and governmental statistics, the development of crop classification methodologies is crucial (Liao et al., 2020), especially for the diversely irrigated agricultural system dominated by those economic crops (Zhong et al., 2019). Therefore, cropland management and crop growth monitoring necessitate the timely acquisition of annual crop maps. These map products assist insurance companies in evaluating disaster-related losses and determining compensation for farmers, offering a more efficient alternative to traditional field visits that are time-consuming and labour-intensive (Chauhan et al., 2020). Bei'an, a county-level city located in Northeast China, primarily focuses on the cultivation of major crops like spring soybean and maize. Additionally, it produces minority crops such as spring wheat. Notably, it serves as a primary production hub for these major crops during the summer season in Northeast China. Distinguishing between soybean and maize from neighbouring crops is challenging due to their similar crop phenology and spectral

profiles (Wang et al., 2019a). This similarity often leads to time-consuming visual interpretation for crop mapping tasks. Addressing the misclassification issue arising from similar phenological stages, spectral features and cropping patterns remains a challenge. Existing solutions seek to enhance feature learning that extracts spectral or polarimetric information effectively between crop classes, using multi-source remote sensing data combined with state-of-the-art classification algorithms to improve crop classification performance.

Traditional (shallow) machine learning models have demonstrated their inherent capabilities in mapping crop types and their spatial distribution, even with limited training datasets (Debats et al, 2016). Typical algorithms such as Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Networks (ANNs), and Decision Trees have been widely employed for crop mapping (Erinjery et al., 2018; Inglada et al., 2015; Löw et al., 2013; Immitzer et al., 2012). While these models efficiently learn from extracted data features, surpassing traditional algorithms in handling dimensional and complex data spaces, they often require feature engineering in remote sensing of vegetation. This involves feature selection of input variables to eliminate redundancies and extract variables that best represent the desired response variable. Such engineering often requires prior knowledge with domain expertise to derive meaningful features from the data, which may struggle when faced with unknown systems exhibiting intricate interactions between objects within the imagery context. In contrast, deep learning models, characterized by their layered convolutional neural structures, excel in end-to-end feature learning. Their architectures, comprising numerous functional layers and transformations, efficiently extract features from high-dimensional remote sensing imagery, unveiling complex and hierarchical relationships within data (Kattenborn et al., 2021).

Synthetic Aperture Radar (SAR) and Polarimetric SAR (PolSAR) data are extensively used for classification tasks due to their resilience against atmospheric and illumination conditions and their ability to penetrate cloud cover. SAR backscatter is especially sensitive to vegetation's three-dimensional structure structure and therefore has a great potential for differentiating crop types and monitoring crop growth (Kattenborn et al., 2021; Qu et al., 2020, Liao et al., 2020). Polarimetric parameters, derived from target decomposition algorithms, have been highlighted for their capacity in image classification (Liao et al., 2020; He et al., 2020; Xie et al., 2019; Gao et al., 2018; Fang et al., 2018). Additionally, interferometric coherence, which measures the correlation between the phases of two complex SAR images acquired at different times,

provides complementary information to SAR intensity. This attribute makes it particularly sensitive and valuable for monitoring land cover changes (Mohammadimanesh et al., 2018), and crop monitoring (Nasirzadehdizaji et al., 2021). Zhang et al. (2017) and Shang et al. (2019) also demonstrated the use of both phase and amplitude information in SAR image classification through CNNs. However, the potential of SAR-derived features in differentiating crops with overlapping growth stages and similar spectral signatures has yet to be fully explored. While Wang et al. (2019a) investigated the enhancement of multispectral features to improve the separability between soybean and maize in Bei'an, the complexity of cropping patterns in these areas still presents significant challenges for accurate crop identification. To address these gaps, there is a need to design tailored classification frameworks that can optimally leverage the utility of SAR signal characteristics.

In addition, integrating SAR data with other remote sensing sources - often referred to as data fusion - can offer richer information and multi-dimensional perspectives that allow for a more holistic understanding of the observed region, thereby enhancing classification performance. While studies combining multitemporal SAR data with multispectral data have shown promising results in enhancing the performance of crop classification through deep learning methods (e.g. Van Tricht et al., 2018; Liao et al., 2020; Adrian et al., 2021), there's potential to further refine these methods by integrating the phenological stages of specific crops. Given that each phenological stage of a crop may present distinct representations, whether through scattered signals or reflectance, capturing these nuances accurately can greatly enhance the ability to differentiate between crop types (Bargiel, 2017). This highlights the importance of using remote sensing techniques and advanced classification algorithms that can discern these subtle yet crucial variations across different growth stages of crops in time series data.

From a deep learning perspective, there has been a growing interest in multi-model networks. These networks, where they are either jointly connected for ensemble learning or combined in a hybrid architecture, have been employed for tasks like land cover classification, human activity recognition, and yield prediction (Barbosa et al., 2020; Branson et al., 2018; Lottes et al., 2018; Zhang et al., 2018; Hamad et al., 2020; Shendryk et al., 2019). This provides opportunities to explore the viability of such combined use of different networks in crop classification for specific regions, especially considering the utilisation of multi-source remote sensing data. Furthermore, it's also worth evaluating the transferability of these pre-trained networks in the context of crop mapping. Specifically, it is important to assess whether these

pre-trained networks can be effectively applied across geospatially various regions and across different years. Additionally, providing visual interpretations of the learning process for hidden features within deep learning models can provide insights into their operational mechanisms and enhance users' understanding of their decision-making processes (Xu et al. 2021).

1.5 Aim, Objectives and Research Questions

The aim of this research is to develop approaches for accurate crop mapping at multiple scales for Bei'an County, China, by employing deep learning models that utilize multi-temporal and multi-source remote sensing datasets. This contributes to identifying and understanding the distribution of non-irrigated croplands and local cropping patterns. The following objectives were addressed in this research with the associated research questions presented following these:

Objective 1 - to develop a framework for multi-temporal crop mapping in Bei'an County by using polarimetric SAR-derived data combined with deep learning methods (Chapter 3)

- Which SAR-derived features play a key role in identifying specific crops?
- How does the developed model perform in comparison to existing models with respect to crop mapping performance and handling imbalanced class distribution?
- To what extent do crops' phenology impact the performance of in-season crop mapping?
- Is the deep learning approach proposed in this study interpretable?

Objective 2 - to construct a sophisticated deep learning architecture that combines multiple models for county-level crop mapping based on the fusion of multi-temporal optical and SAR datasets for Bei'an County (Chapter 4)

- Does the integration of multispectral imagery with SAR data improve the accuracy of crop classification?
- When compared with other models, how does this hybrid model architecture perform?
- In what ways do data augmentation techniques enhance crop classification accuracy?
- Is it feasible to interpret the model based on the features it has learned?

Objective 3 - to design a deep learning based approach tailored for mapping areas of intercropping in Bei'an using interferometric SAR coherence and high resolution (5m) multispectral data (Chapter 5).

- In the context of mapping specific intercropping patterns, how do the temporal and FCN-based models perform?
- How transferable is the developed method, both interannually and spatially?
- Can the model's learning process be unveiled to interpret the features learning process?

1.6 Thesis Structure

This thesis is structured around seven chapters:

Chapter 1: This chapter offers an introductory overview of the research context, detailing the motivations and general background of the project. It provides a foundational introduction to the application of deep learning in remote sensing, followed by highlighting the existing challenges and research gaps as described in the problem statement.

Chapter 2: A comprehensive review of both traditional and deep learning-based methods in agriculture and crop mapping using SAR and optical imagery is presented. This chapter also discusses the pros and cons of various deep learning models and explores visualization techniques for model interpretation.

Chapter 3: This chapter introduces a joint ensemble learning approach that employs two temporal models to extract multi-temporal features from dual-pol SAR data. The goal is to predict crop types in Bei'an for the year 2017. The proposed Conv1D-LSTM model surpasses existing methods by optimally selecting SAR features (m-chi decomposition parameters) and effectively capturing temporal dependencies throughout the entire crop growth cycle. Additionally, it is also capable of handling data with inherently imbalanced class distribution.

Chapter 4: A novel deep learning framework for multi-temporal crop mapping is proposed, which is based on the fusion of polarimetric features and multispectral reflectance. The introduced 3D-ConvSTAR model, which connects 3D-CNN layers with the convolutional recurrent layers (ConvSTAR), demonstrated enhanced classification performance for crop mapping compared to previously designed architectures.

Chapter 5: This chapter presents an ensemble learning for temporal models based on a threshold-based decision fusion strategy to enhance crop mapping performance, specifically for local intercropping patterns. This method combines two temporal models (transformer and attention-based LSTM) with the Random Forest algorithm using a rule-based combination of probability outputs and investigates the synergistic potential of interferometric SAR coherence and multispectral bands for intercropping classification.

Chapter 6: A comprehensive synthesis of the research findings is presented. This chapter revisits and answers the research questions and discusses the limitations of the study, implications on the results, and recommendations for future research.

Chapter 7: This concluding chapter summarises the main research findings of the thesis, reflecting on the challenges encountered, insights acquired, and the knowledge gained throughout the research.

References

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

Barbosa, A., Trevisan, R., Hovakimyan, N. and Martin, N.F., 2020. Modeling yield response to crop management using convolutional neural networks. *Computers and Electronics in Agriculture*, *170*, p.105197.

Bargiel, D., 2017. A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sensing of Environment*, *198*, pp.369-383.

Belgiu, M. and Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*, *204*, pp.509-523.

Blickensdörfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasmi, S. and 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*, p.112831.

Boulila, W., Ghandorh, H., Khan, M.A., Ahmed, F. and Ahmad, J., 2021. A novel CNN-LSTM-based approach to predict urban expansion. *Ecological Informatics*, *64*, p.101325.

Branson, S., Wegner, J.D., Hall, D., Lang, N., Schindler, K. and Perona, P., 2018. From Google Maps to a fine-grained catalog of street trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, *135*, pp.13-30.

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B. and Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sensing of Environment*, *210*, pp.35-47.

Carfagna, E. and Gallego, F.J., 2005. Using remote sensing for agricultural statistics. *International Statistical Review*, *73*(3), pp.389-404.

UN General Assembly., 2015. Transforming our world: *the 2030 Agenda for Sustainable Development*. United Nations: New York, NY, USA

Chauhan, S., Darvishzadeh, R., Lu, Y., Boschetti, M. and Nelson, A., 2020. Understanding wheat lodging using multi-temporal Sentinel-1 and Sentinel-2 data. *Remote Sensing of Environment*, *243*, p.111804.

Chen, B., Zheng, H., Wang, L., Hellwich, O., Chen, C., Yang, L., Liu, T., Luo, G., Bao, A. and Chen, X., 2022. A joint learning Im-BiLSTM model for incomplete time-series Sentinel-2A data imputation and crop classification. *International Journal of Applied Earth Observation and Geoinformation*, *108*, p.102762.

Chen, S. and Wang, H., 2014, October. SAR target recognition based on deep learning. In *2014 International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 541-547.

Chen, Y., Lu, D., Moran, E., Batistella, M., Dutra, L.V., Sanches, I.D.A., da Silva, R.F.B., Huang, J., Luiz, A.J.B. and de Oliveira, M.A.F., 2018. Mapping croplands, cropping patterns, and crop types using MODIS time-series data. *International Journal of Applied Earth Observation and Geoinformation*, *69*, pp.133-147.

Debats, S.R., Luo, D., Estes, L.D., Fuchs, T.J. and Caylor, K.K., 2016. A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. *Remote Sensing of Environment*, *179*, pp.210-221.

Dong, J., Xiao, X., Menarguez, M.A., Zhang, G., Qin, Y., Thau, D., Biradar, C. and Moore III, B., 2016. Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. *Remote Sensing of Environment*, *185*, pp.142-154.

Erinjery, J.J., Singh, M. and Kent, R., 2018. Mapping and assessment of vegetation types in the tropical rainforests of the Western Ghats using multispectral Sentinel-2 and SAR Sentinel-1 satellite imagery. *Remote Sensing of Environment*, *216*, pp.345-354.

Fang, Y., Zhang, H., Mao, Q. and Li, Z., 2018. Land cover classification with gf-3 polarimetric synthetic aperture radar data by random forest classifier and fast super-pixel segmentation. *Sensors*, *18*(7), p.2014.

FAO, 2015. World programme for the census of agriculture 2020. *Rome: F. a. AO ot U. Nations. Retrieved from FAO Statistical Development Series*.

FAO, F., 2018. The future of food and agriculture: alternative pathways to 2050. *Food and Agriculture Organization of the United Nations Rome*.

Flasiński, M., 2016. *Introduction to artificial intelligence*. Springer.

Fritz, S., See, L., Bayas, J.C.L., Waldner, F., Jacques, D., Becker-Reshef, I., Whitcraft, A., Baruth, B., Bonifacio, R., Crutchfield, J. and Rembold, F., 2019. A comparison of global agricultural monitoring systems and current gaps. *Agricultural Systems*, *168*, pp.258-272.

Gallo, I., Ranghetti, L., Landro, N., La Grassa, R. and Boschetti, M., 2023. In-season and dynamic crop mapping using 3D convolution neural networks and sentinel-2 time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, *195*, pp.335-352.

Gao, H., Wang, C., Wang, G., Zhu, J., Tang, Y., Shen, P. and Zhu, Z., 2018. A crop classification method integrating GF-3 PolSAR and Sentinel-2A optical data in the Dongting Lake Basin. *Sensors*, *18*(9), p.3139.

Guo, S., Lv, X. and Hu, X., 2021. Farmers' land allocation responses to the soybean rejuvenation plan: evidence from “typical farm” in Jilin, China. *China Agricultural Economic Review*, *13*(3), pp.705-719.

Hamad, R.A., Yang, L., Woo, W.L. and Wei, B., 2020. Joint learning of temporal models to handle imbalanced data for human activity recognition. *Applied Sciences*, *10*(15), p.5293.

He, C., He, B., Tu, M., Wang, Y., Qu, T., Wang, D. and Liao, M., 2020. Fully convolutional networks and a manifold graph embedding-based algorithm for polsar image classification. *Remote Sensing*, *12*(9), p.1467.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778.

Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural computation*, *9*(8), pp.1735-1780.

Iizumi, T. and Ramankutty, N., 2015. How do weather and climate influence cropping area and intensity?. *Global Food Security*, *4*, pp.46-50.

Immitzer, M., Atzberger, C. and Koukal, T., 2012. Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. *Remote Sensing*, *4*(9), pp.2661-2693.

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P. and Koetz, B., 2015. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing*, *7*(9), pp.12356-12379.

Ji, S., Zhang, C., Xu, A., Shi, Y. and Duan, Y., 2018. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sensing*, *10*(1), p.75.

Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M. and Lobell, D.B., 2019. Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sensing of Environment*, *228*, pp.115-128.

Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G. and Wang, J., 2018. A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy*, *92*, pp.141-152.

Johnson, D.M., 2019. Using the Landsat archive to map crop cover history across the United States. *Remote Sensing of Environment*, *232*, p.111286.

Kattenborn, T., Leitloff, J., Schiefer, F. and Hinz, S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *173*, pp.24-49.

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M. and Inman, D.J., 2021. 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, *151*, p.107398.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *Nature*, *521*(7553), pp.436-444.

Li, J., Zhang, B. and Huang, X., 2022. A hierarchical category structure based convolutional recurrent neural network (HCS-ConvRNN) for Land-Cover classification using dense MODIS Time-Series data. *International Journal of Applied Earth Observation and Geoinformation*, *108*, p.102744.

Li, X., Xu, F., Lyu, X., Gao, H., Tong, Y., Cai, S., Li, S. and Liu, D., 2021. Dual attention deep fusion semantic segmentation networks of large-scale satellite remote-sensing images. *International Journal of Remote Sensing*, *42*(9), pp.3583-3610.

Liao, C., Wang, J., Xie, Q., Baz, A.A., Huang, X., Shang, J. and He, Y., 2020. Synergistic use of multi-temporal RADARSAT-2 and VENμS data for crop classification based on 1D convolutional neural network. *Remote Sensing*, *12*(5), p.832.

Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. *Field Crops Research*, *143*, pp.56-64.

Lottes, P., Behley, J., Milioto, A. and Stachniss, C., 2018. Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. *IEEE Robotics and Automation Letters*, *3*(4), pp.2870-2877.

Löw, F., Michel, U., Dech, S. and Conrad, C., 2013. Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using support vector machines. *ISPRS Journal of Photogrammetry and Remote Sensing*, *85*, pp.102-119.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G. and Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *152*, pp.166-177.

Maponya, M.G., Van Niekerk, A. and Mashimbye, Z.E., 2020. Pre-harvest classification of crop types using a Sentinel-2 time-series and machine learning. *Computers and Electronics in Agriculture*, *169*, p.105164.

Massey, R., Sankey, T.T., Congalton, R.G., Yadav, K., Thenkabail, P.S., Ozdogan, M. and Meador, A.J.S., 2017. MODIS phenology-derived, multi-year distribution of conterminous US crop types. *Remote Sensing of Environment*, *198*, pp.490-503.

Mäyrä, J., Keski-Saari, S., Kivinen, S., Tanhuanpää, T., Hurskainen, P., Kullberg, P., Poikolainen, L., Viinikka, A., Tuominen, S., Kumpula, T. and Vihervaara, P., 2021. Tree species classification from airborne

hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sensing of Environment*, *256*, p.112322.

Mazzia, V., Khaliq, A. and Chiaberge, M., 2019. 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), p.238.

Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Brisco, B. and Motagh, M., 2018. Multi-temporal, multi-frequency, and multi-polarization coherence and SAR backscatter analysis of wetlands. *ISPRS Journal of Photogrammetry and Remote Sensing*, *142*, pp.78-93.

Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Gill, E. and Molinier, M., 2019. A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem. *ISPRS Journal of Photogrammetry and Remote Sensing*, *151*, pp.223-236.

Mou, L., Bruzzone, L. and Zhu, X.X., 2018. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(2), pp.924-935.

Nasirzadehdizaji, R., Cakir, Z., Sanli, F.B., Abdikan, S., Pepe, A. and Calo, F., 2021. Sentinel-1 interferometric coherence and backscattering analysis for crop monitoring. *Computers and Electronics in Agriculture*, *185*, p.106118.

Ni, R., Tian, J., Li, X., Yin, D., Li, J., Gong, H., Zhang, J., Zhu, L. and Wu, D., 2021. An enhanced pixel-based phenological feature for accurate paddy rice mapping with Sentinel-2 imagery in Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, *178*, pp.282-296.

Oliphant, A.J., Thenkabail, P.S., Teluguntla, P., Xiong, J., Gumma, M.K., Congalton, R.G. and Yadav, K., 2019. Mapping cropland extent of Southeast and Northeast Asia using multi-year time-series Landsat 30-m data using a random forest classifier on the Google Earth Engine Cloud. *International Journal of Applied Earth Observation and Geoinformation*, *81*, pp.110-124.

Portmann, F.T., Siebert, S. and Döll, P., 2010. MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochemical Cycles*, *24*(1).

Pott, L.P., Amado, T.J.C., Schwalbert, R.A., Corassa, G.M. and Ciampitti, I.A., 2021. Satellite-based data fusion crop type classification and mapping in Rio Grande do Sul, Brazil. *ISPRS Journal of Photogrammetry and Remote Sensing*, *176*, pp.196-210.

Qu, Y., Zhao, W., Yuan, Z. and Chen, J., 2020. Crop mapping from sentinel-1 polarimetric time-series with a deep neural network. *Remote Sensing*, *12*(15), p.2493.

Ray, D.K., Mueller, N.D., West, P.C. and Foley, J.A., 2013. Yield trends are insufficient to double global crop production by 2050. *PloS One*, *8*(6), p.e66428.

Rußwurm, M. and Körner, M., 2020. Self-attention for raw optical satellite time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *169*, pp.421-435.

Sharma, A., Liu, X., Yang, X. and Shi, D., 2017. A patch-based convolutional neural network for remote sensing image classification. *Neural Networks*, *95*, pp.19-28.

Shendryk, Y., Rist, Y., Ticehurst, C. and Thorburn, P., 2019. Deep learning for multi-modal classification of cloud, shadow and land cover scenes in PlanetScope and Sentinel-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *157*, pp.124-136.

Shi, M., Currier, K., Liu, Z., Janowicz, K., Wiedemann, N., Verstegen, J., McKenzie, G., Graser, A., Zhu, R. and Mai, G., 2023. Thinking Geographically about AI Sustainability. *AGILE: GIScience Series*, *4*, p.42.

Song, X.P., Potapov, P.V., Krylov, A., King, L., Di Bella, C.M., Hudson, A., Khan, A., Adusei, B., Stehman, S.V. and Hansen, M.C., 2017. National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey. *Remote Sensing of Environment*, *190*, pp.383-395.

Tilman, D., Balzer, C., Hill, J. and Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, *108*(50), pp.20260-20264.

Turkoglu, M.O., D'Aronco, S., Perich, G., Liebisch, F., Streit, C., Schindler, K. and Wegner, J.D., 2021. Crop mapping from image time series: Deep learning with multi-scale label hierarchies. *Remote Sensing of Environment*, *264*, p.112603.

Desa, U.N., 2019. World population prospects 2019: Highlights. *New York (US): United Nations Department for Economic and Social Affairs*, *11*(1), p.125.

Van Tricht, K., Gobin, A., Gilliams, S. and Piccard, I., 2018. Synergistic use of radar Sentinel-1 and optical Sentinel-2 imagery for crop mapping: A case study for Belgium. *Remote Sensing*, *10*(10), p.1642.

VoPham, T., Hart, J.E., Laden, F. and Chiang, Y.Y., 2018. Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. *Environmental Health*, *17*(1), pp.1-6.

Wang, L., Dong, Q., Yang, L., Gao, J. and Liu, J., 2019a. Crop classification based on a novel feature filtering and enhancement method. *Remote Sensing*, *11*(4), p.455.

Wang, L., Zhang, G., Wang, Z., Liu, J., Shang, J. and Liang, L., 2019b. Bibliometric analysis of remote sensing research trend in crop growth monitoring: A case study in China. *Remote Sensing*, *11*(7), p.809.

Wang, S., Azzari, G. and Lobell, D.B., 2019c. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment*, *222*, pp.303-317.

Wang, S., Di Tommaso, S., Deines, J.M. and Lobell, D.B., 2020. Mapping twenty years of corn and soybean across the US Midwest using the Landsat archive. *Scientific Data*, *7*(1), p.307.

Wang, Y., Fang, S., Zhao, L., Huang, X. and Jiang, X., 2022. Parcel-based summer maize mapping and phenology estimation combined using Sentinel-2 and time series Sentinel-1 data. *International Journal of Applied Earth Observation and Geoinformation*, *108*, p.102720.

Wei, P., Chai, D., Lin, T., Tang, C., Du, M. and Huang, J., 2021. Large-scale rice mapping under different years based on time-series Sentinel-1 images using deep semantic segmentation model. *ISPRS Journal of Photogrammetry and Remote Sensing*, *174*, pp.198-214.

Wei, S., Zhang, H., Wang, C., Wang, Y. and Xu, L., 2019. Multi-temporal SAR data large-scale crop mapping based on U-Net model. *Remote Sensing*, *11*(1), p.68.

Wen, Y., Li, X., Mu, H., Zhong, L., Chen, H., Zeng, Y., Miao, S., Su, W., Gong, P., Li, B. and Huang, J., 2022. Mapping corn dynamics using limited but representative samples with adaptive strategies. *ISPRS Journal of Photogrammetry and Remote Sensing*, *190*, pp.252-266.

Werbos, P.J., 1990. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, *78*(10), pp.1550-1560.

Wu, X., Xiao, X., Yang, Z., Wang, J., Steiner, J. and Bajgain, R., 2021. Spatial-temporal dynamics of maize and soybean planted area, harvested area, gross primary production, and grain production in the Contiguous United States during 2008-2018. *Agricultural and Forest Meteorology*, *297*, p.108240.

Xie, Q., Wang, J., Liao, C., Shang, J., Lopez-Sanchez, J.M., Fu, H. and Liu, X., 2019. On the use of Neumann decomposition for crop classification using multi-temporal RADARSAT-2 polarimetric SAR data. *Remote Sensing*, *11*(7), p.776.

Xu, J., Yang, J., Xiong, X., Li, H., Huang, J., Ting, K.C., Ying, Y. and Lin, T., 2021. Towards interpreting multi-temporal deep learning models in crop mapping. *Remote Sensing of Environment*, *264*, p.112599.

Xu, J., Zhu, Y., Zhong, R., Lin, Z., Xu, J., Jiang, H., Huang, J., Li, H. and Lin, T., 2020. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. *Remote Sensing of Environment*, *247*, p.111946.

You, N., Dong, J., Huang, J., Du, G., Zhang, G., He, Y., Yang, T., Di, Y. and Xiao, X., 2021. The 10-m crop type maps in Northeast China during 2017–2019. *Scientific Data*, *8*(1), p.41.

You, N., Dong, J., Li, J., Huang, J. and Jin, Z., 2023. Rapid early-season maize mapping without crop labels. *Remote Sensing of Environment*, *290*, p.113496.

Yang, L., Wang, L., Huang, J., Mansaray, L.R. and Mijiti, R., 2019. Monitoring policy-driven crop area adjustments in northeast China using Landsat-8 imagery. *International Journal of Applied Earth Observation and Geoinformation*, *82*, p.101892.

Zhang, C., Di, L., Hao, P., Yang, Z., Lin, L., Zhao, H. and Guo, L., 2021. Rapid in-season mapping of corn and soybeans using machine-learned trusted pixels from Cropland Data Layer. *International Journal of Applied Earth Observation and Geoinformation*, *102*, p.102374.

Zhang, C., Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J. and Atkinson, P.M., 2018. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *140*, pp.133-144.

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J. and Atkinson, P.M., 2019. Joint Deep Learning for land cover and land use classification. *Remote Sensing of Environment*, *221*, pp.173-187.

Zhang, Z., Wang, H., Xu, F. and Jin, Y.Q., 2017. Complex-valued convolutional neural network and its application in polarimetric SAR image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(12), pp.7177-7188.

Zhao, Z.Q., Zheng, P., Xu, S.T. and Wu, X., 2019. Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(11), pp.3212-3232.

Zhong, L., Gong, P. and Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. *Remote Sensing of Environment*, *140*, pp.1-13.

Zhong, L., Hu, L. and Zhou, H., 2019. Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, *221*, pp.430-443.

Chapter 2 Literature Review

The primary focus of this chapter is to provide a review of the methods, evolutions, and challenges specifically encountered in crop classification tasks. Through this examination, this chapter seeks to identify research gaps and potential areas for future investigation, thereby contributing to the enhancement of deep learning applications in crop mapping. Broader reviews of deep learning in agriculture and vegetation remote sensing, which offer an extensive understanding of application domains, model architectures, implementation details, and model assessments, are presented by Kamilaris et al. (2018) and Kattenborn et al. (2021).

2.1 Traditional Remote-Sensing-Based Crop Mapping

Since the early 1970s, satellite-based remote sensing has been used to monitor cropping area, grain yield, and grain production (Lobell, 2013; Fritz et al., 2019). This field has seen considerable advancements, especially in the domain of crop mapping at large spatial scales. Such progress has been facilitated by the advent of newly available satellite imagery with moderate resolution, coupled with the development of advanced classification algorithms (Zhong et al., 2014; Massey et al., 2017; Cai et al., 2018; Defourny et al., 2019). In operational crop mapping systems, notable examples include the Cropland Data Layer (CDL) products developed by the U.S. Department of Agriculture (USDA) National Agricultural Statistics Service (NASS). These products, generated annually using data from multiple satellite sensors such as Landsat and Sentinel-2, have become a cornerstone in national-scale agricultural monitoring (USDA-NASS, 2022). Similarly, the Annual Crop Inventory (ACI) by Agriculture and Agri-Food Canada (AAFC) leverages satellite images from Landsat-8, Sentinel-2, and RADARSAT-2 sensors. The ACI produces maps at a 30 m spatial resolution on a yearly basis, covering a wide range of crop types (Fisette et al., 2013).

Traditional crop classification in remote sensing focused on the distinctive spectral features of crops. For example, Yang et al. (2011) emphasized the importance of spectral bands, such as the shortwave infrared, to enhance classification accuracy. Similarly, Boryan et al. (2011) described the use of optical bands from various satellites in the supervised classification training for the NASS CDL program. However, the challenge to differentiate different crops, which often have similar spectral characteristics during peak-growing seasons, necessitated the development of more sophisticated approaches. Recent studies developed multi-temporal crop

classification schemes to differentiate crops based on time series information that covers varying growth patterns and planting times of crops since the crops' phenology exhibits strong temporal dependencies between the single images of the multitemporal stack (Skriver, 2011; Foerster et al., 2012; Zhang et al., 2015; Sun et al., 2019). This concept has been combined with the use of statistical models to incorporate temporal dependencies, such as Hidden Markov models (Siachalou et al., 2015; Leite et al., 2011). Additionally, some studies have implemented threshold-based methods, defining phenological metrics from vegetation index (VI) time series, like the interval between VI peaks (Fan et al., 2014) or the timing of maximum VI (Walker et al., 2015). These advanced methods generally employ curve-fitting functions, including linear regression (Funk and Budde, 2009), wavelet transform (Galford et al., 2008), and logistic functions (Gonsamo et al., 2016), to fit pre-defined spectral features.

To reduce the demands for human-designed classification rules, machine learning models have been increasingly utilized for processing time series remote sensing observations to effectively identify crops. For example, Song et al. (2017) implemented a Decision Tree (DT) model for national-scale soybean mapping, while Zhang et al. (2014) used the Support Vector Machine (SVM) to measure maize cultivated areas at a provincial scale. Li et al. (2019) further demonstrated the superiority of object-based SVM over pixel-based methods in achieving higher crop classification accuracy. Notably, most large-scale crop classification studies, especially those focusing on maize and soybean, have employed the Random Forest (RF) algorithm (e.g., Zhong et al., 2014; Pelletier et al., 2016; Zhong et al., 2016; Bargiel, 2017; Wang et al., 2019; Mestre-Quereda et al., 2020; You and Dong, 2020; Hao et al., 2020; You et al., 2021; Zhang et al., 2022; Xia et al., 2022; Wen et al., 2022; Blickensdörfer et al., 2022).

While these machine learning techniques have proven robust in processing high-dimensional datasets and learning complex patterns, they often still rely on manual feature engineering. This process can be time-consuming, labour-intensive, and requires substantial domain knowledge to extract distinctive features from raw data that accurately represent crop growth characteristics (Cai et al., 2018; You and Dong, 2020; Kattenborn et al., 2021). Additionally, these methods may not fully explore the sequential relationship of multi-temporal satellite observations, potentially leading to information loss in time series inputs (Zhong et al., 2019; Xu et al., 2021). Therefore, there is a growing need for more advanced approaches in crop mapping applications that can comprehensively capture multidimensional changes, including

geographical and spectral variations, as well as the temporal dynamics of the agricultural landscape.

2.2 Deep Learning in Agricultural Remote Sensing

Remote sensing has become increasingly essential in modern farming, facilitating the collection of diverse spatiotemporal geoinformation that enhances resource efficiency and minimizes environmental impacts of agriculture (Mulla, 2013). In this context, deep learning has emerged as a promising approach, gaining popularity with its diverse applications (e.g. Kamilaris et al., 2018; Kattenborn et al., 2021). Deep learning is characterized by its 'end-to-end' deep neural networks, which perform hierarchical transformations through multi-scale convolution operators (LeCun et al., 2015; Schmidhuber, 2015), or by exploring temporal dynamics using recurrent units in processing remote sensing image data (Mandic and Chambers, 2001). Moreover, deep learning has been effectively employed in various domains of agricultural remote sensing, such as land cover classification (Kussul et al., 2017; Zhang et al., 2019), weed detection (Milioto et al., 2017; Dyrmann et al., 2017), crop mapping (Zhong et al., 2019; Rußwurm and Körner, 2017), and yield prediction (Yang et al., 2019; Chen et al., 2019). The studies are presented as examples in Table 2-1 to reveal the diverse applications of deep learning in agricultural remote sensing. These studies are not the result of a comprehensive or systematic search but a selection of representative applications and, therefore, might not holistically represent the entire scope of current research in this field.

Table 2-1. Summary of studies using deep learning in agricultural remote sensing.

Application	Sensor	Scale	Approach	Reference
Land use and land cover (LULC) classification	USGS National Map Urban Area Imagery collection (spatial resolution of 0.5 m)	Local	Deep convolutional neural network (DCNN)	(Luus et al., 2015)
	Landsat-8 (spatial resolution of 30 m), Sentinel-1 (spatial resolution of 10 m)	Regional, 28,000 km^2	1D-CNN, 2D-CNN	(Kussul et al., 2017)
	Pléiades (spatial resolution of 2 m)	Regional, 42,000 hectares	Recurrent Neural Networks (RNNs)	(Ienco et al., 2017)
	Landsat-8 (spatial resolution of 30 m)	Regional, 771 km^2	2D-CNN	(Sharma et al., 2017)
	Vexcel UltraCam Xp (spatial resolution of 0.5 m)	Local	Object-based convolutional neural networks (OCNN)	(Zhang et al., 2019)
	RADARSAT-2 (a resolution of 5.2 m in the range direction, 7.6 m in the azimuth direction)	Regional	Fully convolutional network (FCN)	(Mohammadimanesh et al., 2019)
	Sentinel-2 (spatial resolution of 10 m)	Regional, 2,640 km^2	Recurrent-Convolutional Neural Network (R-CNN)	(Mazzia et al., 2020)

Weed Detection	Industrial-level imaging sensors (JAI camera)	Fine	2D-CNN	(Milioto et al., 2017)
	Terrestrial sensor	Fine	Fully convolutional network (FCN)	(Dyrmann et al., 2017)
	DJI Phantom 3 Professional (spatial resolution of 0.01 m)	Local, one hectare	AlexNet	(dos Santos Ferreira et al., 2017)
	Terrestrial, AI AD-130 GE (spatial resolution of 0.0001 m)	Fine	Fully convolutional network (FCN)	(Lottes et al., 2018)
	Terrestrial, digital single-lens reflex (DSLR) camera (Nikon D7200) (spatial resolution of 0.001 m)	Fine	Deep convolutional neural network (DCNN)	(Gao et al., 2020)
Crop mapping	Landsat-8 (spatial resolution of 30 m), Sentinel-1 (spatial resolution of 10 m)	Regional, 28,000 km^2	1D-CNN, 2D-CNN	(Kussul et al., 2017)
	Sentinel-2 (spatial resolution of 10 m)	Regional, 102 km × 42 km	Long short-term memory (LSTM)	(Rußwurm and Körner, 2017)
	Gaofen-2 (spatial resolution of 4 m), Gaofen-1 (spatial resolution of 15 m)	Regional	3D-CNN	(Ji et al., 2018)
	Landsat-8 (spatial resolution of 30 m)	Regional, 25,840 hm^2	2D-CNN	(Zhang et al., 2018)
	Landsat (spatial resolution of 30 m)	Regional, 638,767 acres	Deep Neural Network (DNN)	(Cai et al., 2018)
	Formosat-2	Regional 24 km × 24 km area	Temporal 1D-CNNs (TempCNNs)	(Pelletier et al., 2019)
	Landsat (spatial resolution of 30 m)	Regional	1D-CNN	(Zhong et al., 2019)
	Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR) (range and azimuth pixel spacing of 1.66 m and 1 m), Satellite, RapidEye (spatial resolution of 5 m)	Local	Object-based convolutional neural networks and support vector machine (OSVM-OCNN)	(Li et al., 2019)
	Sentinel-1 (spatial resolution of 10 m)	Regional	U-Net	(Wei et al., 2019)
	Sentinel-1 (spatial resolution of 10 m)	Regional	1D-CNN, long short-term memory (LSTM), gated recurrent unit RNNs (GRU RNNs)	(Zhao et al., 2019)
	Sentinel-1 (spatial resolution of 10 m)	Regional, 1,210 km^2	Deep convolutional neural network (DCNN) and long short-term memory (LSTM)	(Zhou et al., 2019)
	Sentinel-1 (spatial resolution of 10 m)	Local, 254 hectares	A combination of a fully convolutional network (FCN) and a convolutional long short-term memory (ConvLSTM) network	(Teimouri et al., 2019)
	Sentinel-1 (spatial resolution of 10 m)	Local, 10 km × 10 km	Depthwise separable convolution recurrent neural network (DSCRNN)	(Qu et al., 2020)
	Sentinel-2 (spatial resolution of 10 m)	Regional	Hybrid convolutional neural network-random forest (CNN-RF)	(Yang et al., 2020)
	Landsat Analysis Ready Data (ARD) (spatial resolution of 30 m)	Regional	Attention-based bidirectional long short-term memory (AtLSTM)	(Xu et al., 2020)

	Sentinel-2 (spatial resolution of 10 m)	Regional	Transformer	(Rußwurm amd Körner, 2020)
	RADARSAT-2, VENµS (spatial resolution of 10 m)	Local	1D-CNN	(Liao et al., 2020)
	Sentinel-2 (spatial resolution of 10 m)	Regional	Attention-based convolutional neural network (CNN)	(Wang et al., 2021)
	Sentinel-1 (spatial resolution of 10 m)	Regional	U-Net	(Wei et al., 2021)
	Sentinel-1 and Sentinel-2 (spatial resolution of 10 m)	Local, 254.3 hectares	3D U-Net	(Adrian et al., 2021)
	Sentinel-2 (spatial resolution of 10 m)	Regional	Convolutional recurrent neural network (ConvSTAR)	(Turkoglu et al., 2021)
	Sentinel-1 and Sentinel-2 (spatial resolution of 10 m)	Regional	3D-CNN	(Teimouri et al., 2022)
	Sentinel-2 (spatial resolution of 10 m)	Regional	3D-CNN	(Gallo et al., 2023)
	Sentinel-1, Sentinel-2 (spatial resolution of 10 m), PlansetScope (spatial resolution of 3 m)	Local	Recurrent Neural Networks (RNNs)	(Rußwurm et al., 2023)
Yield prediction	Airborne, spatial resolution of 0.04 – 0.2 m	Local, 160 hectares	2D-CNN	(Yang et al., 2019)
	DJI Phantom 4 Pro (spatial resolution of 0.00016 m)	Local, 67 m × 6 m	Faster region-based convolutional neural network (R-CNN)	(Chen et al., 2019)
	Moderate Resolution Imaging Spectroradiometer (MODIS) (spatial resolution of 500 m)	National	3D-CNN and attention-based LSTM	(Nejad et al., 2022)

Data pre-processing and preparation for satellite or aerial imagery in agricultural applications, such as calibration, polarimetric decompositions (Liao et al., 2020), atmospheric correction (Rußwurm and Körner, 2017), image segmentation (Li et al., 2019), denoising (Adrian et al., 2021), and feature selection (Yang et al., 2020), often pose significant time-consuming challenges. However, advancements in deep learning models offer promising solutions. For instance, Rußwurm and Körner (2020) demonstrated that models incorporating recurrence and self-attention mechanisms yield higher classification accuracy on raw and cloudy Sentinel 2 data than conventional convolutional-based approaches.

Advanced deep learning architectures, including variants of RNNs (Turkoglu et al., 2021), attention-based networks (Xu et al., 2021), and ensemble learning (Dou et al., 2021), have been developed to enhance crop classification accuracy. Additionally, combining these models with data fusion techniques (Teimouri et al., 2022; Van Tricht et al., 2018; Tao et al., 2022) and data augmentation methods (Mäyrä et al., 2021; Dimitrovski et al., 2023) can significantly improve model generalizability in real-world prediction scenarios. Additionally, the use of specific satellite constellations, such as the PlanetScope constellation from Planet Labs, could

potentially offer rich spatial detail to facilitate global-scale vegetation assessments using deep learning approaches (Kattenborn et al., 2021). This technological advancement contributes to the Sustainable Development Goals (SDGs) by addressing challenges outlined in the United Nations (UN) agenda, particularly in meeting the growing global demand for food production (Persello et al., 2022).

2.2 Advances in Crop Mapping with Deep Learning

2.2.1 Convolutional neural networks

In recent years, the versatility and specialized architecture of data-driven deep learning networks, due to their end-to-end learning paradigm that enables hierarchical feature representations, have led to increased popularity in remote sensing for image classification tasks. In vegetation remote sensing, Convolutional Neural Networks (CNNs) have emerged as reliable feature extractors. These networks can be categorized into 1D-, 2D-, and 3D-CNNs based on the kernels performing convolutional computations across spatial, temporal, and spectral dimensions (Kattenborn et al., 2021). In crop mapping applications, various studies have applied different CNN architectures. For example, Zhong et al. (2019) developed a 1D-CNN-based architecture incorporating an inception module for multi-scale feature extraction focused on the temporal dimension. Similarly, Liao et al. (2020) utilized 1D-CNN to analyse sequential dependencies within satellite data for crop classification. Dou et al. (2021) proposed an innovative ensemble learning framework combining the 1D-CNN-based networks of Zhong et al. (2019) and Pelletier et al. (2019), respectively, achieving higher crop classification accuracy compared to their standalone versions. However, the primary application of 1D convolution, which processes pixel-level features on either temporal or spectral dimensions, does not explicitly consider their spatial relationships. This limitation has led to the application of 2D-CNNs, which are adept at extracting spatial features from the width and height dimensions of images. In remote sensing, 2D-CNNs are often employed for patch-based multidimensional image classification (Sharma et al., 2017). Kussul et al. (2017) showed that 2D convolution with spatial context achieved slightly higher accuracy in crop classification than 1D convolution only considering the spectral context. Conversely, Zhong et al. (2019) found that a 2D-CNN-based architecture yielded slightly inferior results in crop classification compared to 1D temporal convolution, particularly when dealing with dense satellite image datasets.

While 3D CNNs are rarely used compared to 2D CNNs in remote sensing, they offer a unique advantage in exploiting the relationships between multidimensions; the 3D kernels in the network 'slide' across spatial, temporal and spectral dimensions simultaneously, offering a more comprehensive feature extraction (Gallo et al., 2021). Ji et al. (2018) demonstrated the advantage of 3D convolution over 2D convolution as a feature extractor for spatiotemporal remote sensing data in crop classification. Additionally, Teimouri et al. (2022) proposed a novel 3D-CNN architecture for processing multi-temporal, multi-source satellite data, specifically for crop classification. Most of these studies structure CNNs with fully connected (FC) layers at the end of the architecture, where a single pixel or image patch of input tensors finally corresponds to a target class label. However, another notable CNN-based architecture is the fully convolutional network (FCN), adapted for semantic segmentation tasks (Long et al., 2015). A primary example of this is the U-Net architecture, an encoder-decoder network that progressively compresses input data into a compact representation through the encoder, capturing high-level features. The decoder then reconstructs the data back to its original space, making it suitable for detailed, pixel-wise tasks like image segmentation. This approach has been applied in various studies using either 2D U-Net (e.g., Wei et al., 2019; Zhou et al., 2019; He et al., 2020; Wei et al., 2021) or 3D U-Net (e.g., Adrian et al., 2021; Gallo et al., 2023) for crop mapping at both local and large scales. It is important to note that computational resources are subject to the increasing dimensions of the predictor's structure (spatial, temporal, spectral) and the levels of model complexity (Kattenborn et al., 2021). Therefore, optimizing the model architecture and selecting appropriate input features are essential to enhance the computational efficiency of deep learning methods in remote sensing.

2.2.2 Recurrent neural networks

Recurrent Neural Networks (RNNs) are particularly adept at analysing sequential correlations in remote sensing data, making them suitable for end-to-end analysis of long-term sequence signals from crops across their phenological phases (Zhong et al., 2019; Kattenborn et al., 2021). In satellite-based land cover classification, Rußwurm and Körner (2017) employed an RNN with Long Short-Term Memory (LSTM) to encode sequential dependencies from Sentinel-2 temporal observations for crop type classification. Subsequently, Rußwurm and Körner (2018) enhanced this approach by introducing convolutional LSTM (ConvLSTM) and Gated Recurrent Units (GRUs) to encode both temporal and spatial dependencies in the same dataset. The advantage of LSTM-based models in comparison with other methods is notable

as temporal feature extractors. For instance, Rußwurm and Körner (2017) achieved a classification accuracy of 90.6% using an LSTM model, which outperformed CNN (89.2%) and SVM (40.9%). Further advancing this field, Turkoglu et al. (2021) developed a variant of LSTM, the STAR unit and its convolutional version (ConvSTAR), for crop classification. This approach showed a significant increase in F1 score (23.2%) and overall accuracy (3.9%) compared to the LSTM configuration by Rußwurm and Körner (2017) under the dataset with an imbalanced crop class distribution. Moreover, the integration of attention principles in RNNs has further enhanced their performance. Xu et al. (2020) employed an attention-based LSTM (AtLSTM) in a bidirectional manner to discern temporal patterns in multitemporal satellite data for discriminating maize and soybean. Additionally, Rußwurm and Körner (2020) adapted the Transformer architecture (Vaswani et al., 2017) for processing time-series satellite data, employing this self-attention mechanism on crop classification. Xu et al. (2021) then compared AtLSTM with the Transformer, demonstrating that attention-based methods improved performance over traditional RNN-based approaches. These developments highlight the potential of advanced RNNs and attention-based methods in learning essential sequential dependencies from multi-temporal remote sensing observations, providing opportunities for more sophisticated and accurate approaches used for crop mapping.

2.2.3 Ensemble learning

The optimization and innovative design of deep learning structures, particularly through the integration of CNNs and RNNs, have shown promising advancements in multi-temporal crop mapping. For instance, Mazzia et al. (2019) demonstrated an innovative approach that leverages the spatial pattern recognition capabilities of CNNs combined with the temporal data processing strengths of RNNs. In their method, they concatenated pixel-wise branches of RNNs to capture temporal dynamics, followed by the application of a CNN for satellite-based land cover classification. Rustowicz et al. (2019) employed a similar strategy, where satellite images are first processed by a CNN to extract per-image features, which are then analysed for temporal dependencies using RNNs, effectively integrating spatial and temporal data analysis. Furthermore, Interdonato et al. (2019) developed a dual-branch architecture combining CNNs and RNNs to utilize their complementary strengths, thereby deriving a more comprehensive representation of land cover classification. In general, different architectures can be connected in various ways, such as in a hybrid manner (Roy et al., 2019) or through joint connections (Turkoglu et al., 2021). These approaches imply strategic applications of model ensemble

learning, utilizing the strengths of multiple models to improve overall performance and enhance generalization capabilities. Moreover, integrating CNNs with traditional machine learning models, such as SVM and RF, has also proven beneficial. Studies by Li et al. (2019) and Yang et al. (2020) have shown that such synergistic use of different modelling approaches can significantly enhance classification and prediction accuracy in crop mapping.

2.3 SAR-Based Crop Mapping

As discussed previously, the use of multi-temporal remote sensing data and understanding of crop phenology are crucial in crop mapping. However, optical-based remote sensing methods often face challenges, such as missing data and occlusions, particularly in areas with frequent cloud cover or limited visibility. Additionally, the dependency of optical sensors on weather conditions can hinder the continuous collection of satellite acquisitions, thus impacting the accuracy of crop recognition (Qi et al., 2012; Singha et al., 2019). To address these challenges, Synthetic Aperture Radar (SAR) technology, an active sensing approach, is increasingly being adopted in agricultural applications. SAR is particularly advantageous for large-scale crop mapping due to its ability to penetrate cloud cover, operate independently of daylight conditions, and maintain a regular revisiting frequency (Sonobe et al., 2019; Bargiel, 2017; Xie et al., 2019; Qu et al., 2020; Wei et al., 2021). SAR imaging systems provide detailed information, including the amplitude and phase of received backscatter and polarimetric data, which are essential in capturing the seasonal pattern differences between various crops. The examples of SAR-driven approaches in crop mapping are further detailed in Table 2-2, which presents relevant studies employing SAR inputs and models.

In Table 2-2, Xie et al. (2019) and Sonobe et al. (2019) evaluated crop classification performance using different polarimetric decomposition algorithms applied to quad-pol or dual-pol data. Additionally, Qu et al. (2020) showed that covariance matrix vectors significantly contribute to enhanced crop classification accuracy. Mestre-Quereda et al. (2020) reported that interferometric SAR (InSAR) coherence, particularly from image pairs with a 6-day interval, achieved higher crop classification accuracy compared to a 12-day intervals. In the context of crop growth monitoring, Mandle et al. (2020) investigated the correlation between the dual-pol radar vegetation index (DpRVI) and crop biophysical variables, while Nasirzadehdizaji et al. (2021) used InSAR coherence to represent crops' growth patterns. These SAR features have been proven effective in characterizing a variety of crops, including maize,

soybean, and wheat. However, the application of deep learning methods with SAR features other than backscatter remains relatively rare (Wei et al., 2019; Zhao et al., 2019; Zhou et al., 2019; Teimouri et al., 2019; Wei et al., 2021). Furthermore, the unique imaging geometry of SAR, which significantly differs from optical cameras, results in effects such as layover and displacement of moving objects. This dependence of object appearance on viewing geometry presents substantial challenges for feature learning in deep learning approaches when dealing with complex SAR-related data (Persello et al., 2022). For a comprehensive synthesis of SAR-based applications using deep learning, see Zhu et al. (2021).

Table 2-2. SAR-based crop mapping and crop growth monitoring studies.

Sensor	SAR feature	Method	Reference
Sentinel-1	Backscatter (VV, VH)	RF	(Bargiel, 2017)
RADARSAT-2	Neumann decomposition parameters	RF	(Xie et al., 2019)
TerraSAR-X	m-chi decomposition parameters	Multiple kernel learning (MKL)	(Sonobe et al., 2019)
Sentinel-1	Backscatter (VV, VH)	U-Net	(Wei et al., 2019)
Sentinel-1	Backscatter (VH)	RF	(Singha et al., 2019)
Sentinel-1	Backscatter (VV, VH)	1D-CNN, LSTM, GRU, RF	(Zhao et al., 2019)
Sentinel-1	Backscatter (VV, VH, VH/VV)	DCNN-LSTM	(Zhou et al., 2019)
Sentinel-1	Backscatter (VV, VH)	FCN-LSTM	(Teimouri et al., 2019)
Sentinel-1	Dual-pol radar vegetation index (DpRVI)	Linear regression	(Mandle e al., 2020)
Sentinel-1	Interferometric coherence	RF	(Mestre-Quereda et al., 2020)
Sentinel-1	Covariance parameters	Depthwise separable convolution recurrent neural network (DSCRNN)	(Qu et al., 2020)
Sentinel-1	Interferometric coherence	Mean backscatter analysis	(Nasirzadehdizaji et al., 2021)
Sentinel-1	Backscatter (VV, VH)	U-Net	(Wei et al., 2021)

2.4 SAR-Optical Data Fusion in Crop Mapping

SAR and optical data, each with their distinct imaging geometries and information, exhibit different sensitivities towards crop properties. This distinction presents a unique opportunity for data and information fusion, using these complementary data sources in a synergistic manner to enhance crop mapping (Schmitt and Zhu, 2016; Veloso et al., 2017; Gao et al., 2018). The fusion of SAR and optical data can address the inherent limitations of each technology, combining their strengths for a more comprehensive understanding of crop characteristics. The significant challenge in utilizing optical data for crop mapping is the inconsistency of multi-

temporal acquisitions due to cloud cover. This uncertainty can be effectively mitigated by integrating dense multi-source data, including SAR, which is less affected by atmospheric conditions. The addition of SAR data provides a valuable alternative when optical data is compromised, ensuring continuous and reliable monitoring of crop growth and changes (Pott et al., 2021; Onojeghuo et al., 2023).

In the context of crop mapping, one prevalent approach to data fusion involves aggregating data from multiple sources into a compatible tensor during the preprocessing phase. For instance, Gao et al. (2018) demonstrated this by stacking a single acquisition of polarimetric Gaofen-3 data with a single acquisition of Sentinel-2 derivatives using the principal component analysis (PCA) algorithm, creating a fused vector input for the SVM model. Similarly, Liao et al. (2020) integrated multi-temporal polarimetric RADARSAT-2 data with VENμS multispectral data for temporal models like 1D-CNN and LSTM.

Furthermore, studies such as those by Kussul et al. (2018), Van Tricht et al. (2018), You and Dong (2020), Adrian et al. (2021), Blickensdörfer et al. (2022), and Onojeghuo et al. (2023) have explored data fusion scenarios combining multi-temporal Sentinel-1 and Sentinel-2 data to improve classification accuracies for crop types. While multispectral data generally outperforms SAR data in differentiating between crops, the addition of multi-temporal SAR data tends to increase model performance. However, in scenarios where more Sentinel-2 data were combined with fewer Sentinel-1 data, the Sentinel-1 backscatter (VV and VH) showed limited contribution to early season crop mapping (You and Dong, 2020). Stacking multi-source datasets is subject to significant loss of original information due to the process of normalizing these datasets to a common tensor. This is particularly problematic when considering the viewing geometries and acquisition modes, which may not be directly compatible in terms of combined SAR and optical data (Kattenborn et al., 2021). This highlights the need for careful consideration of data fusion techniques to preserve the integrity of the original data while maximizing the benefits of combined analysis.

Another innovative data fusion technique in crop mapping involves adapting the model architecture to perform fusion through intermediate learned features. In this approach, different data sources are processed in parallel branches of a multi-stream network to extract feature maps. Subsequently, these learned features are concatenated at later stages in the network to form a fused feature representation. For instance, Teimouri et al. (2022) designed a 3D-CNN

architecture with two individual paths to extract spatial-temporal-spectral and spatial-temporal-polarization features from multi-temporal Sentinel-2 and Sentinel-1 images, respectively. Similarly, Tao et al. (2022) proposed a dual-branch parallel U-Net architecture. In their approach, multispectral data are split into visible and invisible bands, with the resulting feature maps from each level of the encoder (desampling) concatenated level by level through the decoder (upsampling), thereby obtaining fused semantic segmentation results. Moreover, Barbosa et al. (2020) conducted a comparative analysis of feature-level concatenation from multi-stream approaches with stacked datasets in the context of crop yield response to site-specific management, using diverse inputs like remote sensing data, elevation, and in-situ maps. Their findings indicated that the highest performance was achieved with concatenation after fully connected layers. In contrast, stacking all predictors before modelling proved to be the least effective method, likely due to the complexity of interrelations among varied input datasets.

A more straightforward approach to data fusion in crop mapping uses merging predictions from multiple models, each tailored to specific datasets. This technique can be implemented using methods such as majority voting, where the most common prediction across models is selected as the final output (Baeta et al., 2017). Alternatively, probabilistic techniques like Conditional Random Fields can be employed to merge predictions by considering the conditional dependencies between them (Branson et al., 2018). Additionally, rule-based decision fusion is another viable method, wherein the output probabilities, also known as ‘soft outputs’ from each model, are combined based on predefined thresholds (Li et al., 2019). However, it is important to note that this data fusion approach primarily focuses on combining outputs in the final decision space, rather than integrating the hidden feature representations from different data sources. As a result, it does not fully exploit the synergistic potential of different sources.

2.5 Interpretation of Deep Learning Models

The interpretation of deep learning models, particularly in relation to crop growth characteristics, is a key aspect of evaluating the reliability of crop mapping methodologies (Hu et al., 2019; Zhong et al., 2019; Xu et al., 2021). Deep learning models, and CNNs in particular, are often labelled as 'black boxes' due to the perceived difficulty in understanding their decision-making processes (Reichstein et al., 2019). However, the architecture of CNNs, mostly characterized by a linear and mostly consecutive sequence of repetitive convolutional

operations followed by basic functions like pooling or activation, can facilitate the interpretation of these models. Despite the complexity introduced by the abundance of parameters, these structural properties can convert abstract vectors into interpretable information and help us understand the internal processes of CNNs. Kattenborn et al. (2021) suggest that the interpretation of CNNs can be approached from two perspectives: feature visualization and feature attribution. Feature visualization involves revealing synthetic outputs derived from intermediate layers of the network, commonly referred to as feature maps. This approach helps in visualizing how different layers of the network respond to specific inputs. On the other hand, feature attribution focuses on identifying which specific feature classes in the data activate the network in a salient manner, thereby providing insights into which aspects of the input data are most influential in the network's decision-making process.

For feature maps, the functionality of individual convolutions in deep learning models can be explained using gradient-ascent approaches, as detailed by Schiefer et al. (2020). In this method, gradient ascent is used to modify an input image in a way that maximizes the activation of a network or a specific layer within it. This process seeks to identify local maxima, whereby the output pixel values are adjusted to maximize activation. The resulting layers, therefore, serve as a reflection of the patterns that the network has discerned as crucial during the training process, essentially revealing what the network has learned to recognize. However, a limitation of this approach lies in the nature of the feature maps it produces. These synthetic outputs are often abstract and unnatural, making it challenging to correlate them with real-world class features in remote sensing data (Kattenborn et al., 2021). Furthermore, while feature visualization can generally inform about the general behaviour of the deep learning model and relevant patterns its layers present, it falls short in distinguishing how these patterns differ among various classes in the data. In contrast, feature attribution offers a more intuitive and traceable method of analysing the learning behaviours of deep learning models, as it is directly based on the input data.

Feature attribution in deep learning models often results in the production of activation maps, which typically represent how the input data activates individual feature layers within the networks. These maps are generated by forward propagating individual input images or pixels through a trained network. For example, Mohammadimanesh et al. (2019) utilized activation maps of a CNN to visualize the characteristic backscatter features of different SAR polarisations for wetland classification. Guidici and Clark (2017) analysed activation maps

from the convolutional layer to interpret what the classifier learned about the spectral dimension, focusing on local spectral regions activated by the layer to discern the distinctive spectral characteristics of various classes. Similarly, Rußwurm and Körner (2018) visualized LSTM cell gate activations to demonstrate how information is aggregated over a sequence. However, a limitation of activation maps is that they are inherently input-specific and not output-specific. This means that while they can show how certain parts or features of the input data trigger activations in the network, they do not directly indicate how these activations contribute to the final class prediction decision. In other words, they do not clarify the relationship between specific input features and the network's outputs and how the input features cause the neurons in the network to activate or respond. While visualization of learned features and weights can provide insights into what the network has learned to recognize, activation visualization offers a more practical view of how the network applies these learned capabilities to specific inputs, thus complementing the overall understanding of the model's functioning.

Gradient Weighted Class Activation Mapping (Grad-CAM) offers an output-specific approach to deep learning interpretation by leveraging class-specific gradients from a classification decision to highlight relevant regions in the input data (Selvaraju et al., 2019). This technique enables a more targeted understanding of which areas in the input are significant for the model's decision-making process. Additionally, gradient backpropagation in deep learning interpretation can indicate the importance of input features for neural network models and particularly provide key temporal features in crop mapping. For instance, Zhong et al. (2019) applied multi-level, one-dimensional convolutional layers to visualize temporal patterns throughout crop growth seasons, linking critical temporal features to crop phenology using deconvolution and guided back-propagation (Zeiler and Fergus, 2014). This method allowed for the recognition of patterns indicative of crop seasonality. Similarly, Rußwurm and Körner (2020) and Xu et al. (2021) employed gradients with respect to input time series from models like AtLSTM and Transformer to identify key growth periods and observation bands, which are critical for interpreting deep learning-based multi-temporal crop mapping approaches. Furthermore, Xu et al. (2021) provided a comprehensive, multi-perspective interpretation of the feature learning pipeline inside deep neural networks. Their approach included analysis of gradients and hidden features to monitor the decisive process of dynamic crop mapping. They also applied soft output analysis to assess the model's confidence in its final classification results. These methodologies not only enhance the interpretability of deep learning models in

crop mapping but also contribute to verifying the accuracy and reliability of the predictions made by deep learning models.

Although the interpretation methods for deep learning models are well-established in most fields of vegetation remote sensing, their application in SAR-related crop mapping is still relatively unexplored. This gap is particularly noteworthy considering the growing number of crop mapping studies that utilize features extracted from time-series satellite observations, including multi-temporal SAR-derived data. There is a need to assess and understand the effects of sequential learning models in this context. The development and application of insights derived from artificial intelligence have the potential to significantly enhance our expertise in various technical areas. This includes a deeper understanding of biophysical and ecological aspects of crop mapping, as well as improving our understanding of the correlations between remote sensing data and the characteristics of different crops. The advancement in this area is not just a technological imperative but also a step towards a more comprehensive understanding of agricultural landscapes. Therefore, a focused effort on leveraging and interpreting AI-driven insights in crop mapping based on multi-source data is essential to propel the field forward and explore new dimensions in our understanding of crop dynamics and environmental interactions.

References

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

Baeta, R., Nogueira, K., Menotti, D. and dos Santos, J.A., 2017, October. Learning deep features on multiple scales for coffee crop recognition. In *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 262-268.

Barbosa, A., Trevisan, R., Hovakimyan, N. and Martin, N.F., 2020. Modeling yield response to crop management using convolutional neural networks. *Computers and Electronics in Agriculture*, *170*, p.105197.

Bargiel, D., 2017. A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sensing of Environment*, *198*, pp.369-383.

Blickensdörfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasmi, S. and 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*, p.112831.

Boryan, C., Yang, Z., Mueller, R. and Craig, M., 2011. Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto International*, *26*(5), pp.341-358.

Branson, S., Wegner, J.D., Hall, D., Lang, N., Schindler, K. and Perona, P., 2018. From Google Maps to a fine-grained catalog of street trees. *ISPRS Journal of Photogrammetry and Remote Sensing*, *135*, pp.13-30.

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B. and Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sensing of Environment*, *210*, pp.35-47.

Chen, Y., Lee, W.S., Gan, H., Peres, N., Fraisse, C., Zhang, Y. and He, Y., 2019. Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages. *Remote Sensing*, *11*(13), p.1584.

Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaute, T. and Savinaud, M., 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*, pp.551-568.

Dimitrovski, I., Kitanovski, I., Kocev, D. and Simidjievski, N., 2023. Current trends in deep learning for Earth Observation: An open-source benchmark arena for image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *197*, pp.18-35.

dos Santos Ferreira, A., Freitas, D.M., da Silva, G.G., Pistori, H. and Folhes, M.T., 2017. Weed detection in soybean crops using ConvNets. *Computers and Electronics in Agriculture*, *143*, pp.314-324.

Dou, P., Shen, H., Li, Z. and Guan, X., 2021. Time series remote sensing image classification framework using combination of deep learning and multiple classifiers system. *International Journal of Applied Earth Observation and Geoinformation*, *103*, p.102477.

Dyrmann, M., Jørgensen, R.N. and Midtiby, H.S., 2017. RoboWeedSupport-Detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network. *Advances in Animal Biosciences*, *8*(2), pp.842-847.

Fan, C., Zheng, B., Myint, S.W. and Aggarwal, R., 2014. Characterizing changes in cropping patterns using sequential Landsat imagery: An adaptive threshold approach and application to Phoenix, Arizona. *International Journal of Remote Sensing*, *35*(20), pp.7263-7278.

Fisette, T., Rollin, P., Aly, Z., Campbell, L., Daneshfar, B., Filyer, P., Smith, A., Davidson, A., Shang, J. and Jarvis, I., 2013, August. AAFC annual crop inventory. In *2013 Second International Conference on Agro-Geoinformatics*, pp. 270-274

Foerster, S., Kaden, K., Foerster, M. and Itzerott, S., 2012. Crop type mapping using spectral–temporal profiles and phenological information. *Computers and Electronics in Agriculture*, *89*, pp.30-40.

Fritz, S., See, L., Bayas, J.C.L., Waldner, F., Jacques, D., Becker-Reshef, I., Whitcraft, A., Baruth, B., Bonifacio, R., Crutchfield, J. and Rembold, F., 2019. A comparison of global agricultural monitoring systems and current gaps. *Agricultural Systems*, *168*, pp.258-272.

Funk, C. and Budde, M.E., 2009. Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe. *Remote Sensing of Environment*, *113*(1), pp.115-125.

Galford, G.L., Mustard, J.F., Melillo, J., Gendrin, A., Cerri, C.C. and Cerri, C.E., 2008. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. *Remote Sensing of Environment*, *112*(2), pp.576-587.

Gallo, I., La Grassa, R., Landro, N. and Boschetti, M., 2021. Sentinel 2 Time Series Analysis with 3D Feature Pyramid Network and Time Domain Class Activation Intervals for Crop Mapping. *ISPRS International Journal of Geo-Information*, *10*(7), p.483.

Gallo, I., Ranghetti, L., Landro, N., La Grassa, R. and Boschetti, M., 2023. In-season and dynamic crop mapping using 3D convolution neural networks and sentinel-2 time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, *195*, pp.335-352.

Gao, H., Wang, C., Wang, G., Zhu, J., Tang, Y., Shen, P. and Zhu, Z., 2018. A crop classification method integrating GF-3 PolSAR and Sentinel-2A optical data in the Dongting Lake Basin. *Sensors*, *18*(9), p.3139.

Gao, J., French, A.P., Pound, M.P., He, Y., Pridmore, T.P. and Pieters, J.G., 2020. Deep convolutional neural networks for image-based Convolvulus sepium detection in sugar beet fields. *Plant Methods*, *16*(1), pp.1-12.

Gonsamo, A. and Chen, J.M., 2016. Circumpolar vegetation dynamics product for global change study. *Remote Sensing of Environment*, *182*, pp.13-26.

Gu, Y., Wang, Y. and Li, Y., 2019. A survey on deep learning-driven remote sensing image scene understanding: Scene classification, scene retrieval and scene-guided object detection. *Applied Sciences*, *9*(10), p.2110.

Guidici, D. and Clark, M.L., 2017. One-Dimensional convolutional neural network land-cover classification of multi-seasonal hyperspectral imagery in the San Francisco Bay Area, California. *Remote Sensing*, *9*(6), p.629.

Hao, P., Di, L., Zhang, C. and Guo, L., 2020. Transfer Learning for Crop classification with Cropland Data Layer data (CDL) as training samples. *Science of The Total Environment*, *733*, p.138869.

Hartling, S., Sagan, V., Sidike, P., Maimaitijiang, M. and Carron, J., 2019. Urban tree species classification using a WorldView-2/3 and LiDAR data fusion approach and deep learning. *Sensors*, *19*(6), p.1284.

He, C., He, B., Tu, M., Wang, Y., Qu, T., Wang, D. and Liao, M., 2020. Fully convolutional networks and a manifold graph embedding-based algorithm for polsar image classification. *Remote Sensing*, *12*(9), p.1467.

Hu, Q., Sulla-Menashe, D., Xu, B., Yin, H., Tang, H., Yang, P. and Wu, W., 2019. A phenology-based spectral and temporal feature selection method for crop mapping from satellite time series. *International Journal of Applied Earth Observation and Geoinformation*, *80*, pp.218-229.

Ienco, D., Gaetano, R., Dupaquier, C. and Maurel, P., 2017. Land cover classification via multitemporal spatial data by deep recurrent neural networks. *IEEE Geoscience and Remote Sensing Letters*, *14*(10), pp.1685-1689.

Interdonato, R., Ienco, D., Gaetano, R. and Ose, K., 2019. DuPLO: A DUal view Point deep Learning architecture for time series classificatiOn. *ISPRS Journal of Photogrammetry and Remote Sensing*, *149*, pp.91-104.

Ji, S., Zhang, C., Xu, A., Shi, Y. and Duan, Y., 2018. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sensing*, *10*(1), p.75.

Kamilaris, A. and Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, pp.70-90.

Kattenborn, T., Leitloff, J., Schiefer, F. and Hinz, S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *173*, pp.24-49.

Kussul, N., Lavreniuk, M., Skakun, S. and Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, *14*(5), pp.778-782.

Kussul, N., Mykola, L., Shelestov, A. and Skakun, S., 2018. Crop inventory at regional scale in Ukraine: developing in season and end of season crop maps with multi-temporal optical and SAR satellite imagery. *European Journal of Remote Sensing*, *51*(1), pp.627-636.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *Nature*, *521*(7553), pp.436-444.

Leite, P.B.C., Feitosa, R.Q., Formaggio, A.R., da Costa, G.A.O.P., Pakzad, K. and Sanches, I.D.A., 2011. Hidden Markov Models for crop recognition in remote sensing image sequences. *Pattern Recognition Letters*, *32*(1), pp.19-26.

Li, H., Zhang, C., Zhang, S. and Atkinson, P.M., 2019. A hybrid OSVM-OCNN method for crop classification from fine spatial resolution remotely sensed imagery. *Remote Sensing*, *11*(20), p.2370.

Li, Y., Huang, X. and Liu, H., 2017. Unsupervised deep feature learning for urban village detection from high-resolution remote sensing images. *Photogrammetric Engineering & Remote Sensing*, *83*(8), pp.567-579.

Liao, C., Wang, J., Xie, Q., Baz, A.A., Huang, X., Shang, J. and He, Y., 2020. Synergistic use of multi-temporal RADARSAT-2 and VENμS data for crop classification based on 1D convolutional neural network. *Remote Sensing*, *12*(5), p.832.

Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. *Field Crops Research*, *143*, pp.56-64.

Long, J., Shelhamer, E. and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440.

Lottes, P., Behley, J., Milioto, A. and Stachniss, C., 2018. Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. *IEEE Robotics and Automation Letters*, *3*(4), pp.2870-2877.

Mandal, D., Kumar, V., Ratha, D., Dey, S., Bhattacharya, A., Lopez-Sanchez, J.M., McNairn, H. and Rao, Y.S., 2020. Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data. *Remote Sensing of Environment*, *247*, p.111954.

Mandic, D.P. and Chambers, J., 2001. *Recurrent neural networks for prediction: learning algorithms, architectures and stability*. *John Wiley & Sons*, Inc..

Massey, R., Sankey, T.T., Congalton, R.G., Yadav, K., Thenkabail, P.S., Ozdogan, M. and Meador, A.J.S., 2017. MODIS phenology-derived, multi-year distribution of conterminous US crop types. *Remote Sensing of Environment*, *198*, pp.490-503.

Mäyrä, J., Keski-Saari, S., Kivinen, S., Tanhuanpää, T., Hurskainen, P., Kullberg, P., Poikolainen, L., Viinikka, A., Tuominen, S., Kumpula, T. and Vihervaara, P., 2021. Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sensing of Environment*, *256*, p.112322.

Mazzia, V., Khaliq, A. and 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), p.238.

Mestre-Quereda, A., Lopez-Sanchez, J.M., Vicente-Guijalba, F., Jacob, A.W. and Engdahl, M.E., 2020. Time-series of Sentinel-1 interferometric coherence and backscatter for crop-type mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, pp.4070-4084.

Milioto, A., Lottes, P. and Stachniss, C., 2017. Real-time blob-wise sugar beets vs weeds classification for monitoring fields using convolutional neural networks. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*, pp.41-48.

Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Gill, E. and Molinier, M., 2019. A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem. *ISPRS Journal of Photogrammetry and Remote Sensing*, *151*, pp.223-236.

Mulla, D.J., 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, *114*(4), pp.358-371.

Nasirzadehdizaji, R., Cakir, Z., Sanli, F.B., Abdikan, S., Pepe, A. and Calo, F., 2021. Sentinel-1 interferometric coherence and backscattering analysis for crop monitoring. *Computers and Electronics in Agriculture*, *185*, p.106118.

Nejad, S.M.M., Abbasi-Moghadam, D., Sharifi, A., Farmonov, N., Amankulova, K. and Lászlź, M., 2022. Multispectral crop yield prediction using 3D-convolutional neural networks and attention convolutional LSTM approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, pp.254-266.

Onojeghuo, A.O., Miao, Y. and Blackburn, G.A., 2023. Deep ResU-Net Convolutional Neural Networks Segmentation for Smallholder Paddy Rice Mapping Using Sentinel 1 SAR and Sentinel 2 Optical Imagery. *Remote Sensing*, *15*(6), p.1517.

Pelletier, C., Webb, G.I. and Petitjean, F., 2019. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, *11*(5), p.523.

Persello, C., Wegner, J.D., Hänsch, R., Tuia, D., Ghamisi, P., Koeva, M. and Camps-Valls, G., 2022. Deep learning and earth observation to support the sustainable development goals: Current approaches, open challenges, and future opportunities. *IEEE Geoscience and Remote Sensing Magazine*, *10*(2), pp.172-200.

Pott, L.P., Amado, T.J.C., Schwalbert, R.A., Corassa, G.M. and Ciampitti, I.A., 2021. Satellite-based data fusion crop type classification and mapping in Rio Grande do Sul, Brazil. *ISPRS Journal of Photogrammetry and Remote Sensing*, *176*, pp.196-210.

Qi, Z., Yeh, A.G.O., Li, X. and Lin, Z., 2012. A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data. *Remote Sensing of Environment*, *118*, pp.21-39.

Qu, Y., Zhao, W., Yuan, Z. and Chen, J., 2020. Crop mapping from sentinel-1 polarimetric time-series with a deep neural network. *Remote Sensing*, *12*(15), p.2493.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. and Prabhat, F., 2019. Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), pp.195-204.

Roy, S.K., Krishna, G., Dubey, S.R. and Chaudhuri, B.B., 2019. HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, *17*(2), pp.277-281.

Rustowicz, R., Cheong, R., Wang, L., Ermon, S., Burke, M., Lobell, D., 2019. Semantic segmentation of crop type in africa: A novel dataset and analysis of deep learning methods. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 75-82.

Rußwurm, M. and Körner, M., 2017. Multi-temporal land cover classification with long short-term memory neural networks. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, pp.551-558.

Rußwurm, M. and Körner, M., 2020. Self-attention for raw optical satellite time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *169*, pp.421-435.

Rußwurm, M. and Körner, M., 2018. Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS International Journal of Geo-Information*, *7*(4), p.129.

Rußwurm, M., Courty, N., Emonet, R., Lefèvre, S., Tuia, D. and Tavenard, R., 2023. End-to-end learned early classification of time series for in-season crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, *196*, pp.445-456.

Schiefer, F., Kattenborn, T., Frick, A., Frey, J., Schall, P., Koch, B. and Schmidtlein, S., 2020. Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, *170*, pp.205-215.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. *Neural Networks*, *61*, pp.85-117.

Schmitt, M. and Zhu, X.X., 2016. Data fusion and remote sensing: An ever-growing relationship. *IEEE Geoscience and Remote Sensing Magazine*, *4*(4), pp.6-23.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618-626.

Sharma, A., Liu, X., Yang, X. and Shi, D., 2017. A patch-based convolutional neural network for remote sensing image classification. *Neural Networks*, *95*, pp.19-28.

Siachalou, S., Mallinis, G. and Tsakiri-Strati, M., 2015. A hidden Markov models approach for crop classification: Linking crop phenology to time series of multi-sensor remote sensing data. *Remote Sensing*, *7*(4), pp.3633-3650.

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

Skriver, H., 2011. Crop classification by multitemporal C-and L-band single-and dual-polarization and fully polarimetric SAR. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(6), pp.2138-2149.

Sonobe, R., 2019. Parcel-based crop classification using multi-temporal TerraSAR-X dual polarimetric data. *Remote Sensing*, *11*(10), p.1148.

Sun, C., Bian, Y., Zhou, T. and Pan, J., 2019. Using of multi-source and multi-temporal remote sensing data improves crop-type mapping in the subtropical agriculture region. *Sensors*, *19*(10), p.2401.

Tao, C., Meng, Y., Li, J., Yang, B., Hu, F., Li, Y., Cui, C. and Zhang, W., 2022. MSNet: multispectral semantic segmentation network for remote sensing images. *GIScience & Remote Sensing*, *59*(1), pp.1177-1198.

Teimouri, M., Mokhtarzade, M., Baghdadi, N. and Heipke, C., 2022. Fusion of time-series optical and SAR images using 3D convolutional neural networks for crop classification. *Geocarto International*, *37*(27), pp.15143-15160.

Teimouri, N., Dyrmann, M. and Jørgensen, R.N., 2019. A novel spatio-temporal FCN-LSTM network for recognizing various crop types using multi-temporal radar images. *Remote Sensing*, *11*(8), p.990.

Turkoglu, M.O., D'Aronco, S., Perich, G., Liebisch, F., Streit, C., Schindler, K. and Wegner, J.D., 2021. Crop mapping from image time series: Deep learning with multi-scale label hierarchies. *Remote Sensing of Environment*, *264*, p.112603.

USDA-NASS, C. D. L., 2022. USDA National Agricultural Statistics Service Cropland Data Layer. Published crop-specific data layer. Available at: (https://nassgeodata.gmu.edu/CropScape/).

Van Tricht, K., Gobin, A., Gilliams, S. and Piccard, I., 2018. Synergistic use of radar Sentinel-1 and optical Sentinel-2 imagery for crop mapping: A case study for Belgium. *Remote Sensing*, *10*(10), p.1642.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

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

Walker, J.J., De Beurs, K.M. and Wynne, R.H., 2014. Dryland vegetation phenology across an elevation gradient in Arizona, USA, investigated with fused MODIS and Landsat data. *Remote Sensing of Environment*, *144*, pp.85-97.

Wang, S., Azzari, G. and Lobell, D.B., 2019. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment*, *222*, pp.303-317.

Wang, Y., Zhang, Z., Feng, L., Ma, Y. and Du, Q., 2021. A new attention-based CNN approach for crop mapping using time series Sentinel-2 images. *Computers and Electronics in Agriculture*, *184*, p.106090.

Wei, P., Chai, D., Lin, T., Tang, C., Du, M. and Huang, J., 2021. Large-scale rice mapping under different years based on time-series Sentinel-1 images using deep semantic segmentation model. *ISPRS Journal of Photogrammetry and Remote Sensing*, *174*, pp.198-214.

Wei, S., Zhang, H., Wang, C., Wang, Y. and Xu, L., 2019. Multi-temporal SAR data large-scale crop mapping based on U-Net model. *Remote Sensing*, *11*(1), p.68.

Wen, Y., Li, X., Mu, H., Zhong, L., Chen, H., Zeng, Y., Miao, S., Su, W., Gong, P., Li, B. and Huang, J., 2022. Mapping corn dynamics using limited but representative samples with adaptive strategies. *ISPRS Journal of Photogrammetry and Remote Sensing*, *190*, pp.252-266.

Xia, T., He, Z., Cai, Z., Wang, C., Wang, W., Wang, J., Hu, Q. and Song, Q., 2022. Exploring the potential of Chinese GF-6 images for crop mapping in regions with complex agricultural landscapes. *International Journal of Applied Earth Observation and Geoinformation*, *107*, p.102702.

Xie, Q., Wang, J., Liao, C., Shang, J., Lopez-Sanchez, J.M., Fu, H. and Liu, X., 2019. On the use of Neumann decomposition for crop classification using multi-temporal RADARSAT-2 polarimetric SAR data. *Remote Sensing*, *11*(7), p.776.

Xu, J., Yang, J., Xiong, X., Li, H., Huang, J., Ting, K.C., Ying, Y. and Lin, T., 2021. Towards interpreting multi-temporal deep learning models in crop mapping. *Remote Sensing of Environment*, *264*, p.112599.

Xu, J., Zhu, Y., Zhong, R., Lin, Z., Xu, J., Jiang, H., Huang, J., Li, H. and Lin, T., 2020. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. *Remote Sensing of Environment*, *247*, p.111946.

Yalcin, H., 2017, August. Plant phenology recognition using deep learning: Deep-Pheno. In *2017 6th International Conference on Agro-Geoinformatics*, pp. 1-5.

Yang, C., Everitt, J.H. and Murden, D., 2011. Evaluating high resolution SPOT 5 satellite imagery for crop identification. *Computers and Electronics in Agriculture*, *75*(2), pp.347-354.

Yang, Q., Shi, L., Han, J., Zha, Y. and Zhu, P., 2019. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crops Research*, *235*, pp.142-153.

Yang, S., Gu, L., Li, X., Jiang, T. and Ren, R., 2020. Crop classification method based on optimal feature selection and hybrid CNN-RF networks for multi-temporal remote sensing imagery. *Remote Sensing*, *12*(19), p.3119.

You, N. and Dong, J., 2020. Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, *161*, pp.109-123.

You, N., Dong, J., Huang, J., Du, G., Zhang, G., He, Y., Yang, T., Di, Y. and Xiao, X., 2021. The 10-m crop type maps in Northeast China during 2017–2019. *Scientific Data*, *8*(1), p.41.

Zeiler, M.D. and Fergus, R., 2014. Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*, pp.818-833.

Zhang, C., Di, L., Lin, L., Li, H., Guo, L., Yang, Z., Eugene, G.Y., Di, Y. and Yang, A., 2022. Towards automation of in-season crop type mapping using spatiotemporal crop information and remote sensing data. *Agricultural Systems*, *201*, p.103462.

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J. and Atkinson, P.M., 2019. Joint Deep Learning for land cover and land use classification. *Remote Sensing of Environment*, *221*, pp.173-187.

Zhang, G., Xiao, X., Dong, J., Kou, W., Jin, C., Qin, Y., Zhou, Y., Wang, J., Menarguez, M.A. and Biradar, C., 2015. Mapping paddy rice planting areas through time series analysis of MODIS land surface temperature and vegetation index data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *106*, pp.157-171.

Zhang, J., Feng, L. and Yao, F., 2014. Improved maize cultivated area estimation over a large scale combining MODIS–EVI time series data and crop phenological information. *ISPRS Journal of Photogrammetry and Remote Sensing*, *94*, pp.102-113.

Zhang, M., Lin, H., Wang, G., Sun, H. and Fu, J., 2018. Mapping paddy rice using a convolutional neural network (CNN) with Landsat 8 datasets in the Dongting Lake Area, China. *Remote Sensing*, *10*(11), p.1840.

Zhao, H., Chen, Z., Jiang, H., Jing, W., Sun, L. and Feng, M., 2019. Evaluation of three deep learning models for early crop classification using sentinel-1A imagery time series—A case study in Zhanjiang, China. *Remote Sensing*, *11*(22), p.2673.

Zhong, L., Gong, P. and Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. *Remote Sensing of Environment*, *140*, pp.1-13.

Zhong, L., Hu, L. and Zhou, H., 2019. Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, *221*, pp.430-443.

Zhong, L., Yu, L., Li, X., Hu, L. and Gong, P., 2016. Rapid corn and soybean mapping in US Corn Belt and neighboring areas. *Scientific Reports*, *6*(1), p.36240.

Zhou, Y.N., Luo, J., Feng, L. and Zhou, X., 2019. DCN-based spatial features for improving parcel-based crop classification using high-resolution optical images and multi-temporal SAR data. *Remote Sensing*, *11*(13), p.1619.

Zhu, X.X., Montazeri, S., Ali, M., Hua, Y., Wang, Y., Mou, L., Shi, Y., Xu, F. and Bamler, R., 2021. Deep learning meets SAR: Concepts, models, pitfalls, and perspectives. *IEEE Geoscience and Remote Sensing Magazine*, *9*(4), pp.143-172.

Chapter 3 Enhanced Crop Mapping Using Polarimetric SAR Features and Time Series Deep Learning: A Case Study in Bei'an, China

Abstract

Large-scale crop mapping is essential for decision-makers to evaluate agricultural resource usage and estimate crop yields. Annual crop inventory statistics can provide valuable insights into crop growth monitoring, however, due to the variability in crop growth patterns across different locations and times, the ability to generalise near real-time crop classification over large areas is needed. This study develops deep learning based approaches to map agricultural regions at the county level using multi-temporal Sentinel-1 Synthetic Aperture Radar (SAR) data, specifically evaluating the contribution of SAR-derived input predictors for discriminating both majority and minority crops in Bei'an County, Northeast China. The proposed model architecture amalgamates a one-dimensional convolution (Conv1D) neural network with attention-based Long Short-Term Memory (LSTM) to characterise the crop types exhibiting phenological similarities using a range of SAR-derived input predictors. The results are compared with alternative multi-temporal deep learning frameworks, including standalone Conv1D and Transformer models, as well as the machine learning algorithm Random Forest (RF), which serves as the baseline for comparison. The designed architecture (Conv1D-LSTM) achieved the highest F1 scores (maize: 87%, soybean: 86% and other crops: 85%) when applied to an inherently imbalanced dataset, using m-chi decomposition features as input predictors. The results provide superior performance in terms of effectiveness and efficiency compared to other selected models. The monthly in-season crop classification underscores the importance of temporal dependencies and the availability of multi-temporal observations for learning dynamic growth patterns over large areas. Moreover, the interpretation of model learning processes and outcomes is explained through visualising weight distributions and hidden features. This study offers a comprehensive evaluation of essential SAR features in multi-temporal satellite data for accurate crop mapping, utilising advanced deep learning techniques.

Keywords: Crop mapping; Deep learning; Multi-temporal satellite data; Polarimetric Synthetic Aperture Radar (PolSAR); Conv1D; Attention-based LSTM

3.1 Introduction

Accurate crop mapping is essential for dynamically monitoring agricultural productivity, assessing food availability, and supporting decision-making at regional and national levels (Gómez et al., 2016; Mercier et al., 2020; Wu et al., 2021; Blickensdörfer et al., 2022). For instance, early-season crop classification contributes to numerous agricultural applications, such as cropland management, supply chain frameworks, crop insurance, and area-based subsidies supported by governments (Cai et al., 2018). Accurate and near real-time crop classification is greatly needed for understanding changes in growth patterns of specific crops and assessing their socio-economic impacts.

Modern satellite remote sensing technology enables the detection, mapping, and monitoring of agronomic information over various spatial and temporal scales, and can help optimize agricultural practices, reducing the environmental impact of food production and minimizing waste, and ultimately contributing to more sustainable food systems (Thenkabail et al., 2012, Benos et al., 2021). Satellite remote sensing offers opportunities to obtain multi-dimensional data at high spatial and temporal resolutions over large regions. Satellite-based workflows have been widely employed to enhance crop mapping accuracies and produce crop maps at various scales (Azzari and Lobell, 2017). For example, Cropland Data Layer (CDL) and Crop Inventory (CI) are two national-scale, medium-resolution crop map products that are updated annually. The 2022 CDL, a geo-referenced raster layer of the United States Department of Agriculture (USDA), offers geospatial information pertaining to crop types, land cover, and land use at a 30-meter resolution across the United States (USDA-NASS, 2022). In contrast, CI is developed by Agriculture and Agri-Food Canada (AAFC) and provides crop type information for Canada at a 30-meter resolution (Fisette et al., 2013). Generating both products involves the integration of multi-source satellite imagery and comprehensive ground truth data collected via field surveys.

Although the multi-spectral and photosynthetic properties of vegetation can be measured by optical satellite platforms (Veloso et al., 2017), frequent cloud cover can limit the quality of optical acquisitions, significantly impacting the performance of effective differentiation of crops (Sonobe et al., 2014; Griffiths et al., 2019). Methods to reconstruct occluded information, such as linear interpolation (Kandasamy et al., 2013), and probabilistic models like K-nearest neighbour (KNN) imputation and decision trees (Bertsimas et al., 2017), are not only

computationally demanding but also encounter challenges in achieving accurate imputation, particularly for large datasets with significant missing or contaminated data (Khan et al., 2022). To address these limitations, Synthetic Aperture Radar (SAR) sensors, active instruments that can operate independent of illumination and weather conditions, are widely used to retrieve information on the Earth`s surface. While previous studies have demonstrated that time-series SAR features have positive impacts on land cover classification performance (e.g. Ullmann et al., 2014; Zhang et al., 2014), their applications in agricultural domains remain challenging, particularly in comparison to the utilization of optical data (Veloso et al., 2017; Steele-Dunne et al., 2017).

The Sentinel-1 satellite constellation, with the short revisit frequency of six days when both satellites are operational, offers open-access SAR data at C-band with global coverage, ensuring consistent monitoring and mapping applications in agriculture (e.g., McNairn et al., 2009; Inglada et al., 2016; Navarro et al., 2016; Ndikumana et al., 2018; Mullissa al., 2018; Teimouri et al., 2019, Beriaux et al., 2021). However, most studies have focused primarily on radar amplitude information as input predictors. Sentinel-1 can also provide phase information characterised by off-diagonal elements in the scattering matrix, which can improve crop mapping accuracy by investigating the scattering properties of different land cover types (Sun et al., 2019; Qu et al., 2020). Additionally, polarimetric SAR (PolSAR) is an advanced radar remote sensing technique that measures and interprets the polarisation state of backscattered signals, thereby enabling a detailed understanding of the scattering mechanism and enhancing the identification and characterisation of surface features, such as crops. Polarimetric decomposition algorithms have been developed to extract physical information from the target surface, which can break down the random scattering mechanisms into several parameters that can be associated with the target's physical characteristics. The polarimetric parameters are also sensitive to phenological changes in crops (Valcarce-Diñeiro et al., 2018). However, many existing decomposition methods are only applicable to quad-polarisation (quad-pol) sensors (He et al., 2020; Liao et al., 2020; Xie et al., 2019; Gao et al., 2018), which limits their use with dual-polarised platforms like Sentinel-1. Furthermore, quad-polarimetric observations often have reduced swath coverage and limited availability, hindering their application across large areas with high temporal frequency (Sonobe et al., 2019). While dual-polarisation (dual-pol) data has been investigated as an alternative solution for large-scale crop monitoring by several studies (e.g., Heine et al., 2016; Sonobe et al., 2019; Qu et al., 2020; Bhogapurapu et al., 2021;

Hosseini et al., 2022), there remains a pressing need for more refined algorithms that can accurately capture variability both within fields and across landscapes.

Machine learning methods have been widely employed for large-scale agricultural monitoring (e.g., Benos et al., 2021; Li et al., 2020) and crop mapping (e.g., Dong et al., 2018; Xu et al., 2018; Sun et al., 2019; Moumni and Lahrouni, 2021), using satellite remote sensing observations. Other applications include retrieving biophysical features (e.g., Verrelst et al., 2012; Jia et al., 2019) and yield prediction (e.g., Chlingaryan et al., 2018; Baloloy et al., 2018). Commonly used machine learning approaches, such as Random Forest (RF), Decision Tree (DL), and Support Vector Machine (SVM) have demonstrated good performances for multi-temporal crop mapping (Sonobe et al., 2019; King et al., 2017; Xie et al., 2019; Gao et al., 2018; Dong et al., 2018). However, these models typically handle non-temporal data without fully exploring the underlying temporal dependencies, resulting in an incomplete understanding of time-series patterns of input data. Various studies have considered integrating phenological metrics to define crop characteristics (Siachalou et al., 2015; Zhong et al., 2014; Bargiel, 2017), but these empirical features are heavily dependent on expert knowledge and may be specific to local cropping practices.

Recently, deep learning techniques have shown their efficiency and reliability in handling remote sensing data for a range of applications such as land cover classification, crop type classification, plant disease detection, and crop yield prediction (e.g., Kamilaris and Prenafeta-Boldú, 2018; Zhang et al., 2019; Liao et al., 2020, Zhu et al., 2020). These cutting-edge methods employ end-to-end neural network architectures capable of inherently exploring and learning high-dimensional information. Convolutional Neural Networks (CNNs) are extensively employed to extract multi-scale features from remote sensing data (Zhang et al., 2020). Specifically, one-dimensional CNNs (1D-CNNs) have been successfully employed for crop mapping by using one-dimensional convolutional (Conv1D) layers as pixel-based feature extractors processing multi-temporal remote sensing datasets (Zhong et al., 2019; Pelletier et al., 2019; Liao et al., 2020). Additionally, Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Transformers have also demonstrated success in enhancing crop mapping by extracting dynamic temporal features from longer image sequences (Zhou et al., 2017; Zhao et al., 2019; Zhong et al., 2019; Rußwurm and Körner, 2020). In the case of stacked Conv1D layers, the shallow layers extract local features, while deeper layers comprehensively summarize patterns to a larger extent. LSTM units, on the other hand, are designed to memorize

features over long or short-time sequences. It has been shown that deep learning methods are capable of learning hierarchical features across spatiotemporal dimensions and frequently outperform traditional classifiers that rely on manually defined rules for crop classification tasks (Sun et al., 2019; Wei et al., 2019; Zhou et al., 2019).

In this context, there is a clear need for studies that investigate the potential of deep learning approaches and understand the interpretability of these networks, specifically those designed to uncover sequential relationships in multi-temporal remote sensing data for crop mapping. This is particularly important for regions like Northeast China, where obtaining quantitative information from local croplands has been challenging due to annual crop rotation practices. (You et al., 2021). The Northeast region of China has become a significant agricultural region, playing a key role in both domestic agricultural production and international trade, particularly in the cultivation of economically important crops such as soybeans and maize (Dong et al., 2016; Yang et al., 2019). As a result, there is a strong demand for accurate annual crop maps to assist local authorities in establishing near real-time crop monitoring systems for early yield assessments at the county level. However, reliable crop mapping is difficult due to the complex relationship between the location-specific factors (spatial) and the timing-related factors (temporal) that affect crop growth (Qu et al., 2020; Liao et al., 2020; Xu et al., 2020; Xu et al., 2021). Although 1D-CNNs benefit from lower computational complexities (Kiranyaz et al., 2021), their predictive performance depends on the sensitivity of input predictors derived from satellite data and the design of model architectures (Yang et al., 2020; Dou et al., 2021). Crop mapping challenges are further magnified by the limited availability of accurate crop type labels, resulting in imbalanced dataset distributions and reduced crop map accuracies on a large scale, despite the improvements in spatial and temporal resolutions provided by contemporary earth-observing satellites (Wang et al., 2019). Moreover, labour costs, expertise requirements, and the accessibility of multi-temporal remote sensing data continue to pose constraints for large-scale crop monitoring initiatives (Dong et al., 2015; Zhang et al., 2015; You and Dong, 2020; Zhang et al., 2022).

The aim of this study is to develop an approach for multi-temporal crop mapping in Bei'an County in Northeast China using polarimetric SAR-derived data and deep learning. The research addresses the following specific questions: i) which SAR features derived from Sentinel-1 datasets have the greatest impact on crop mapping in Bei'an County? ii) how does the developed Conv1D-LSTM architecture perform in comparison to existing models, such as

Conv1D (Zhong et al, 2019), Conv1D-RF (Yang et al, 2020), Transformer (Rußwurm and Körner, 2020), and the universal baseline model RF, with respect to crop mapping performance and handling imbalanced class distribution? iii) how do phenological factors influence the effectiveness of in-season crop mapping using SAR data? and iv) how is the interpretability of the proposed deep learning approach demonstrated in this study? Specifically, the study develops a joint ensemble learning architecture (Conv1D-LSTM) by combining the attention-based LSTM and Conv1D layers to reduce the recognition error rate stemming from the imbalanced dataset. The developed architecture is compared to existing state-of-the-art deep learning models and the baseline model (RF) for differentiating local major crops such as soybean and maize, which exhibit similar phenological stages. Additionally, we explore various SAR features, including the backscatter coefficient, Grey Level Co-occurrence Matrix (GLCM), covariance matrix parameters, radar vegetation index (RVI), dual-pol RVI (DpRVI), and polarimetric features derived by m-chi decomposition and dual-polarisation entropy/alpha. Due to the potential for high-dimensional SAR features to introduce redundant information, we explored feature selection techniques, such as the Spearman correlation coefficient and the feature importance algorithm (Boruta), to eliminate less relevant features. We then created multiple scenarios to evaluate the importance of specific SAR features in differentiating crops.

3.2 Material and Methods

3.2.1 Study area

Bei'an, located in the transitional zone between the Songnen Plain and the Greater Khingan mountains (47°35'N ~48°33'N, 126°16'E~127°53'E), is a county-level city in the northeast part of Heilongjiang province in China (Figure 3-1). Bei'an spans a total area of 7,149 km^2 and lies within a cold temperate continental monsoon climate zone, with an average annual temperature of 1.2°C, an Effective Accumulated Temperature (EAT) of 18.30 °C - 23.50 °C, a frost-free period of 88 - 120 days, and an average annual precipitation of 529 millimetres. The long summer daylight hours, large diurnal temperature differences in autumn, and rainfall concurrent with the warm season are beneficial for crop growth. This region, also known as one of the world's famous black soil regions, provides enhanced soil fertility, creating improved conditions for agricultural activities, particularly grain cultivation.

Approximately 35.4% of the total area of Bei'an is forested, and cropland occupies around 32.95% of the total area, with spring soybean and maize being the major crops grown (61.8%

and 29.5% of the total sowing area, respectively) (Heihe Social and Economic Statistics Yearbook, 2018). Other minority crop types mainly include spring rice and wheat, along with other land cover types. A network of rivers and canals traverses the area, and the land-use pattern is characterized by a mosaic of agricultural fields interspersed with parcels of forests, grasslands, and wetlands, which contribute to essential irrigation for crop cultivation. According to the local crop sowing scheme, maize is typically sown from late April to late September, whereas soybeans are normally sown from early May to mid-September. These periods might vary annually due to crop rotation cycles over the years in the study area. As illustrated in Figure 3-2, the observed agricultural landscape exhibits a diverse range of field configurations, with large croplands dedicated to the dominant crops (maize and soybean) alongside compact farmlands where multiple crop types are cultivated adjacently. This complex cropping pattern poses potential challenges for satellite-based crop mapping, particularly when relying on medium or low-spatial-resolution imagery.

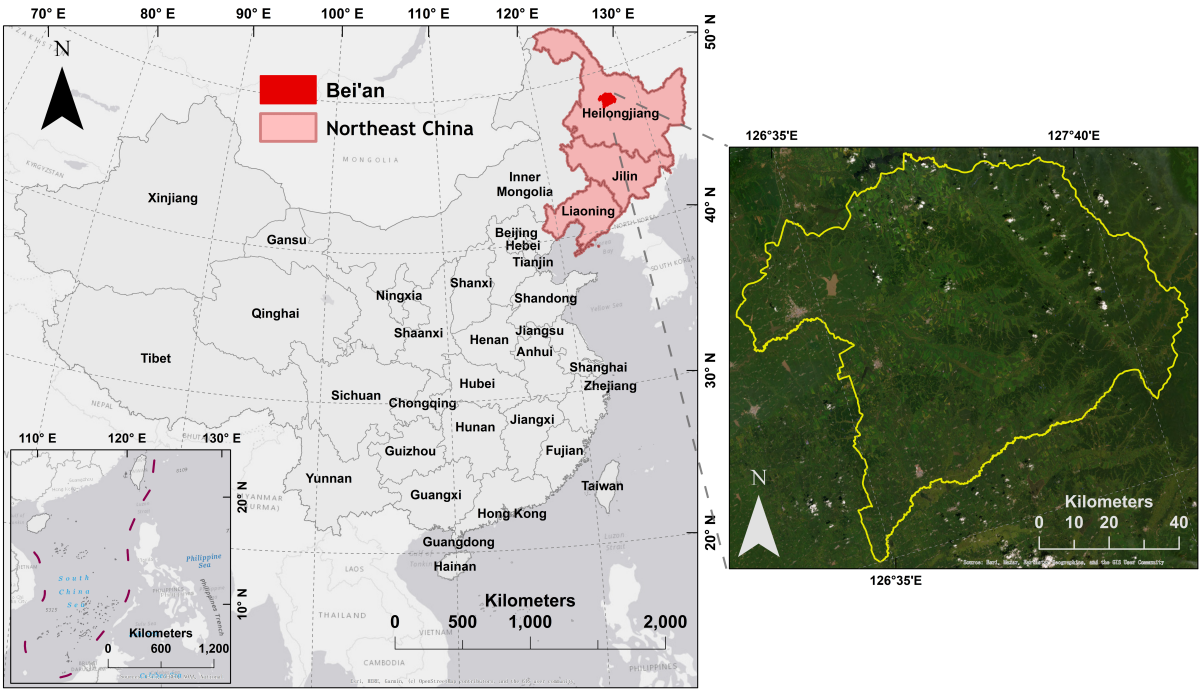

Figure 3-1. The study area of Bei'an County, Northeast China (the right panel shows the county's boundary).

Figure 3-2. Images of maize (left), soybean (centre) and mixed crops (right) were captured in Southeast Bei'an during the crop-growing stage in August 2019.

3.2.2 Ground-truth dataset

During July, August, and September 2017, on-site field surveys were conducted in the study area, covering a total of 21,257 fields to determine crop parcel areas, record crop types, and obtain annual statistics from agricultural household surveys. The surveyed field polygons from these sites were then provided to the Chinese Academy of Agricultural Sciences (CAAS), which subsequently classified the crops using remote sensing imagery combined with the surveyed field polygons. For cropland parcels where mixed crops were intercropped in close proximity, manual digitization and labelling were performed using 5 m spatial resolution RapidEye imagery (Near Infra-red, Red Edge, and Red composite), while Sentinel-2A images (SWIR, Narrow NIR, and Red composite) were applied for delineating larger monocropping cropland parcels. In addition to crops, residential, forest and waterbody areas were also labelled. The reference dataset provided by CAAS contains some crop parcels for rice, wheat, and unidentified crops, which were merged into a single class, namely 'other crop', due to their relatively small sample sizes. The proportion of sample pixels for each class and the field size distribution are illustrated in Figure 3-3. The entire samples, consisting of 3,979,417 labelled pixels, were allocated for training, validation, and testing dataset division using 10-km grids, in line with the data partitioning approach suggested by Zhong et al. (2019), with the respective ratio of 60%, 20%, and 20%. A random selection of 10% (equivalent to 397,942 pixels) from each dataset was made for the purposes of training models in a weakly supervised framework and for conducting feature selection. This means one of the 10% subsets functioned bilaterally as a training set for model development and as input for feature selection processes. The objective of employing feature selection techniques was to identify critical SAR features suitable for classification tasks by eliminating irrelevant features. Following this, the models were trained using the selected features to ensure their relevance and effectiveness. Subsequently, these trained models were evaluated on a separate 10% subset (designated as the

testing dataset) through comparative analysis to determine the most impactful SAR-related input predictors for the task of crop mapping. After determining the most important predictors, the remaining ground-truth labels were incorporated into the 10% dataset, thereby replenishing it for subsequent evaluation of the accuracy improvement achieved by integrating these predictors into the crop mapping process.

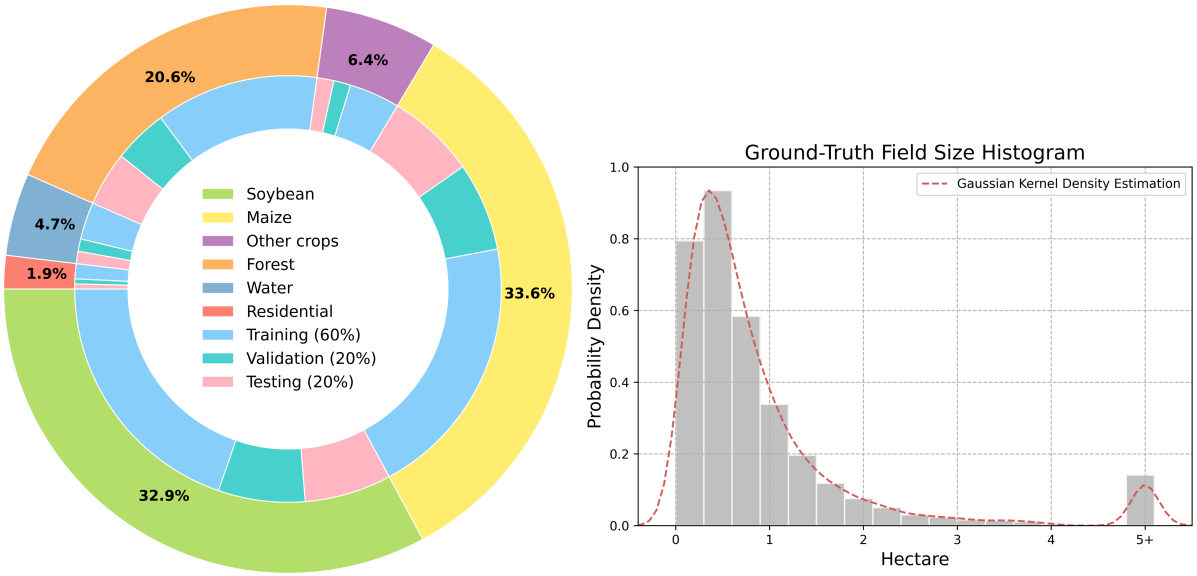

Figure 3-3. The proportion of imbalanced samples for each class (left). The training, validation, and testing datasets are mutually exclusive. Pixels from each class were extracted from these sets in a ratio of 60:20:20. The distribution of crop field size in the Bei'an ground-truth dataset (right). Parcels larger than 4 hectares are aggregated in the last bin of the histogram. The average parcel size is 1.39 hectares.

3.2.3 SAR data collection and pre-processing

Sentinel-1 is a constellation of satellites equipped with C-band SAR sensors. These satellites are positioned 180° apart, with Sentinel-1A launched in April 2014 and Sentinel-1 B launched in April 2016 (anomaly in December 2021 resulted in the end of the mission) by the European Space Agency (ESA). Sentinel-1 Interferometric Wide (IW) Single Look Complex (SLC) products (5 × 20 m spatial resolution) were obtained from the Sentinel-1 Scientific Data Hub (https://scihub.copernicus.eu/). Since the main agricultural practices in the study area typically occur during summer and autumn, data collection focused on the period from early May to late September (May 6th and September 20th). In total, 22 Sentinel-1B acquisitions were collected in 2017, corresponding to the timeframe when the ground-truth dataset was collected. Figure 3-4 shows the monthly distribution of acquisitions, along with the corresponding growth stages of soybean and maize in 2017. The phenological stages for maize and soybean are synthesized in monthly intervals according to Wang et al (2019), with the corresponding crops' appearance

at each phenological stage defined by BBCH-scale codes (Meier et al., 2009). Figure 3-4 provides a detailed illustration of these stages along with their corresponding BBCH codes for each crop type. Soybean reaches full maturity in August and is typically harvested in September. Throughout these two stages, the soybean exhibits a consistent appearance as per the BBCH scale.

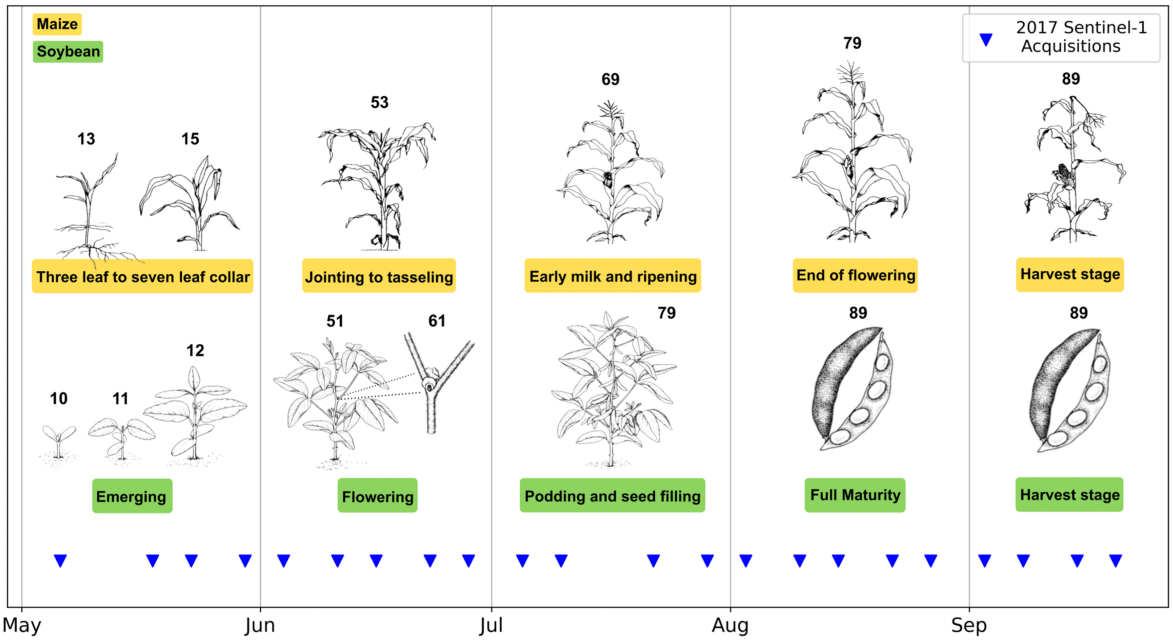

Figure 3-4. The Sentinel-1 acquisitions, obtained from May 6th to September 20th, were analysed alongside the phenological stages for maize and soybean. Note that the numbers at the top of crops' phenological stages are BBCH codes.

The pre-processing of Sentinel-S1B images was conducted using the open-source Sentinel application platform (SNAP 8.0) developed by the European Space Agency (ESA). The pre-processing steps, as suggested by Qu et al (2020) and Li et al (2019), generally include Radiometric Calibration, Polarimetric Matrix Generation, Multilooking, Polarimetric Speckle Filtering, and Geocoding. A 3 x 3 window size adaptive Lee filter was employed to mitigate the impact of speckle noise while preserving information, as suggested by Mahdianpari et al (2017). Intensity data was converted into backscatter coefficient sigma nought (σ^0) in logarithmic dB scale. The cross-ratio of the backscatter is calculated by VH minus VV in terms of the logarithm rules. As the calculation of radar-based vegetation index generally relies on quad-pol data (Kumar et al., 2013), this study employed RVI, a vegetation index specifically tailored for Sentinel-1 backscatter data to monitor crop growth (Nasirzadehdizaji et al., 2019; Tomaszewski et al., 2021).

SAR scattering features and DpRVI were derived from the dual-pol covariance matrix (C_2) using Sentinel-1 Toolbox in SNAP. This 2×2 matrix represents the relationship between the incident field and the scattered field vector, including two real parts (C_{11} and C_{22}) and two complex parts (the real part C_{12} and the imagery part C_{12}). These components of the Sentinel-1 covariance matrix can be directly used as model inputs for crop mapping (Qu et al., 2020). Based on C_2, additional scattering characteristics can be investigated through target decomposition methods to generate further polarimetric parameters correlated with ground objects.

This study evaluated decomposition features derived by dual-polarisation entropy/alpha decomposition (H/α dual-pol decomposition) and m-chi decomposition. All processed images were geo-referenced using the digital elevation model (DEM) from the Shuttle Radar Topography Mission (SRTM 3 arc second) and then resampled to 10-meter spatial resolution. PolSAR pre-processing details can be found in Mandal et al. (2019). A few missing values were imputed by using the KNN algorithm provided by Python's scikit-learn package version 1.2.0, with '4' selected as the optimal number of neighbours (Zhang, 2012). Due to the inconsistent range of values across all considered predictors, a typical Min-Max normalization approach was applied to all input features.

Four main stages, as illustrated in Figure 3-5, were performed in this study's workflow. The Sentinel-1 pre-processing procedures and SAR feature candidates are primarily presented in the subsequent sub-sections of section 3.2.4. Feature selection techniques were applied to eliminate redundant information and assess collinearity between features (see section 3.2.5). Section 3.2.6 provides an overview of the classification approaches. Model implementation was introduced in section 3.2.7. The performance assessment identifies the optimal SAR features and evaluates the generalisability of the proposed method for thematic map prediction and in-season crop classification (section 3.3). Model interpretation was discussed in section 3.4.

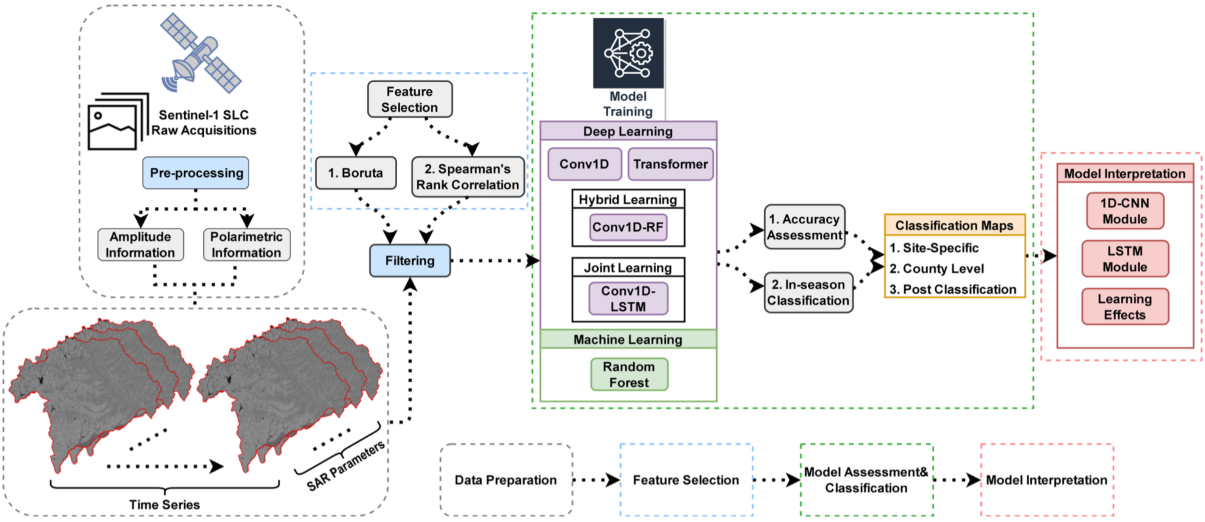

Figure 3-5. The methodology flowchart.

3.2.4 SAR-derived features

3.2.4.1 H/α dual-pol decomposition

Cloude and Pottier (1997) developed an Entropy (H) - Alpha (α) - Anisotropy decomposition technique that was originally calculated through eigenvalues of the 3×3 coherency matrix derived from quad-polarisation data to generate target scattering matrix parameters. Sentinel-1 SLC products provide dual-pol information in covariance matrix with two orthogonal components of the scattered signal. Various studies (e.g. Guo et al., 2018; Ioannidou et al., 2022) have explored the scattering mechanisms of target features for crop classification using modified H/α dual-pol decomposition tailored for dual-polarised Sentinel-1 data, defined as:

$$H = \sum_{i=1}^{2} -P_i log_2 P_i, \ wherein \ P_i = \frac{\lambda_i}{\sum_{i=1}^{2} \lambda_i} \ and \ \lambda_1 \geq \lambda_2 \quad (3\text{-}1)$$

$$\alpha = \sum_{i=1}^{2} P_i \, \alpha_i, \ wherein \ 0 \leq \alpha \leq 90° \quad (3\text{-}2)$$

where λ_1 and λ_2 are eigenvalues of the covariance matrix. The polarimetric scattering entropy H ranging from 0 to 1 is an indicator that measures the degree of scattering randomness. The mean alpha angle α indicates physical scattering characteristics listed in nine zones (Cloude and Pottier, 1997). Based on both parameters, a H/α plane is proposed to represent all random scattering mechanisms and utilised for crop growth analysis (Guo et al, 2018; Salma et al, 2022).

3.2.4.2 m-chi decomposition

The m-chi decomposition is a compact polarimetric decomposition method developed by Raney et al (2012) to analyse scattering characteristics from compact PolSAR data. The π/4 mode of the compact polarimetric SAR system transmits circularly polarised signals and receives orthogonal backscattered signals in linear horizontal and vertical polarisation (Souyris et al., 2005). This approach facilitates extracting scattering information from the target while reducing the number of transmitted and received channels compared to quad-pol systems. Using the covariance matrix constructed from Sentinel-1 data, the m-chi decomposition generates 'pseudo' polarimetric features resembling compact polarimetric parameters. Characterized by four-element Stokes parameters, the degree of polarisation (m_1) is described in Eq. (3-3):

$$m_1 = \frac{\sqrt{S_2^2+S_3^2+S_4^2}}{S_1} \tag{3-3}$$

where $S_{1,2,3,4}$ represents four Stokes parameters for each pixel in the total power over an image field, and they are calculated from the averaged covariance matrix. Additionally, several candidates for a second decomposition parameter can be obtained from the Stokes parameters, such as $Chi\ (\chi)$, a Poincaré variable indicating the degree of ellipticity. The $\sin 2\chi$, also known as the degree of circularity, is expressed in Eq. (3-4):

$$\sin 2\chi = \pm\frac{S_4}{mS_1} \tag{3-4}$$

where transmitted right or left-hand circular polarisation is represented by the positive or negative sign, respectively. Based on these two variables calculated from the Stokes parameters, three parameters for m-chi decomposition can be derived as follows (Eq. (3-5) – Eq. (3-7)):

$$B = \sqrt{\frac{mS_1(1-\sin 2\chi)}{2}} \tag{3-5}$$

$$R = \sqrt{\frac{mS_1(1+\sin 2\chi)}{2}} \tag{3-6}$$

$$G = \sqrt{S_1(1-m)} \tag{3-7}$$

where B, G and R correspond to single-bounce, double-bounce and volume backscattering, respectively.

Previous studies (e.g., Nord et al., 2008; Ainsworth et al., 2009) showed that compact polarimetric data are close to and occasionally equivalent to quad-pol data by which fully polarimetric decompositions are conducted. Although all four-element Stokes parameters can be derived from dual-polarized data like Sentinel-1 SLC products, the challenge lies in the separability between single and double bounce targets due to the ellipticity angle from dual-pol data being close to zero (Raney, 2007). This could reduce the separability between polarimetric scattering types. Despite this challenge, a single scattering type may prevail over agricultural targets during specific growth stages, and the combination of scattering mechanisms alters as the canopy varies (McNairn et al., 2009). Consequently, this study considers complete m-chi decomposition features as input predictors.

3.2.4.3 Dual-pol Radar vegetation index (DpRVI)

The dual-pol radar vegetation index (DpRVI) has previously been demonstrated to exhibit a strong correlation with crop biophysical variables and effectively represents crop growth dynamics (Mandle et al., 2020). The proportion of polarisation of an electromagnetic wave is characterised in terms of the degree of polarisation m_2, as proposed by Barakat (1977) in Eq. (3-8) and Eq. (3-9):

$$m_2 = \sqrt{1 - \frac{4|C_2|}{(Span)^2}}, \quad \in(0, 1) \tag{3-8}$$

$$Span = \lambda_1 + \lambda_2, \quad \lambda_1 \geq \lambda_2 \geq 0 \tag{3-9}$$

where λ_1 and λ_2 are eigenvectors obtained through the eigen-decomposition of the 2×2 covariance matrix C_2 denoted as $\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$, which are normalised with the total power Span. To quantitively assess the scattering strength, Mandle et al (2020) introduced a parameter β, which is subsequently used to derive DpRVI. See Eq. (3-10) and Eq. (3-11):

$$\beta = \frac{\lambda_1}{Span} \tag{3-10}$$

$$DpRVI = 1 - DOP * \beta, \quad \in[0, 1] \tag{3-11}$$

3.2.4.4 GLCM features

The GLCM evaluates the arrangement of grayscale values in an image and is utilized to quantify the variations in texture within images. It calculates spatial features inherent in raster images by determining the probability of occurrence of pixel pairs separated by a predefined distance in a given direction (Szantoi et al., 2013; Zhou et al., 2019). These features are derived from the properties of pixel co-occurrence statistics, calculated within a specified moving window and following certain directions at inter-pixel distances. Moumni and Lahrouni (2021) found that combining SAR backscatter with statistical group features such as GLCM mean, variance and correlation led to better crop classification performance.

3.2.5 Feature selection

The practical model building often encounters limitations when working with modern datasets containing multivariate variables. Many of these variables may be irrelevant to the target classification, making it challenging to determine their correlations. For instance, a large PolSAR image tends to contain data redundancy, especially when generating rich indices from multi-source remote sensing data (Yang et al., 2020; Liao et al., 2020). As a result, extracting relevant features that correlate with ground objects becomes a significant challenge. Feature selection typically considers two aspects, including the multicollinearity between input variables and the correlation between the inputs and the targets. Generally, data redundancy can lead to increased computational costs during model training and may negatively impact performance when the optimal feature combination is uncertain. Hence, feature selection techniques that derive minimal and optimal feature sets are essential for achieving the best possible classification results.

3.2.5.1 Boruta

This study applied one of the wrapper methods designed on the Random Forest algorithm, which is an ensemble method for classification with its unique voting mechanism comprising manifold unbiased decision trees developed independently on multiple bagging samples. It performs a recursive process on multiple feature sets and iteratively eliminates the features less relevant to the label target prediction. The Boruta package (Kursa et al., 2010) was initially implemented in the R environment. This study used an alternative Python-compatible version that leverages Gini impurity to determine feature importance scores, accessed at

https://github.com/scikit-learn-contrib/boruta_py. This method evaluates the correlation between inputs and corresponding label targets.

3.2.5.2 Spearman coefficients

The second step investigates the relevance between multivariant variables by calculating the Spearman rank correlation coefficient. The Spearman rank correlation coefficient evaluates the correlation and strength of the relationship between two variables (Khare et al., 2012). Higher correlation coefficients indicate stronger monotonicity between two variables. In this study, the coefficients from multi-temporal SAR features previously determined by Boruta are assessed. Based on the feature importance analysis provided in the first step, the number of features can be optimally combined for model training. The Spearman correlation coefficient r_s is given by Chambers (1989) in Eq. (3-12):

$$r_s = \frac{\sum_{i=1}^{N}(a_i-\bar{a})(b_i-\bar{b})}{\sqrt{\sum_{i=1}^{N}(a_i-\bar{a})^2}\sqrt{\sum_{i=1}^{N}(b_i-\bar{b})^2}}, \in [-1, 1] \quad (3\text{-}12)$$

where N is the total number of a feature across the whole temporal dimension and a_i, b_i is distinctive features to be measured. The value range of r_s quantifies the correlations and strength between variables, which can be either negatively or positively correlated.

3.2.6 Classification approaches

3.2.6.1 Random Forest

Random Forest (RF), introduced by Breiman (1996), has proven to be a widely used machine learning approach for handling high-dimensional remote sensing data in crop mapping. It is often used as a baseline model to assess model performance based on the comparison within the remote sensing domain (Belgiu and Drăguţ., 2016). RF is an ensemble classifier that operates multiple unbiased decision trees with a bagging strategy to control the prediction variance of the model performance. Given this tree-based random feature sample selection, the problem of overfitting can be minimised (Pal, 2005). However, since RF is not designed to measure time-series dependencies, the input features for RF in this study are only divided into a two-dimensional input shape, consisting of the number of pixel-level features and their corresponding bands. To optimize modelling performance, hyperparameters are tuned to

enhance and maintain the model generalisation. In this study, the optimal number of trees is set to 200 and the minimal number of samples required to split an internal node is set to 4.

3.2.6.2 Conv1D-based architectures

Conv1D is one of the variants of CNN, which employs 1D filters convolving along the temporal dimension to extract patterns over a time-series sequence. Shallow levels of the stacked Conv1D layers concentrate on feature learning at a local scale, while deeper layers summarise learned features holistically, extending to larger extents. However, the final configuration of the CNN layers is initially tested on simple structures and then empirically developed with increased complexity to achieve stable model performance. In this study, two architectures of Conv1D-based networks for crop classification performance comparison are applied, as depicted in Figure 3-6. One of the architectures is taken from Zhong et al (2019), which adopts the inception module introduced by Szegedy et al (2015) to extract multi-temporal patterns at multiple scales using a combination of Conv1D layers and a pooling layer.

The second Conv1D-RF architecture, derived from Yang et al (2020), follows the inception framework but employs a different classifier at the end. This hybrid Conv1D structure replaces the originally applied Softmax classifier with Random Forest in the last fully connected layer, producing final outputs via the unique voting mechanism designed for RF. This CNN-RF hybridization strategy considers the advantages provided by RF against outliers and overfitting (Kwak et al., 2021). The pre-trained Conv1D module outputs a feature vector (512×1) from the first fully connected layer, which is then fed into RF classifier for the final classification.

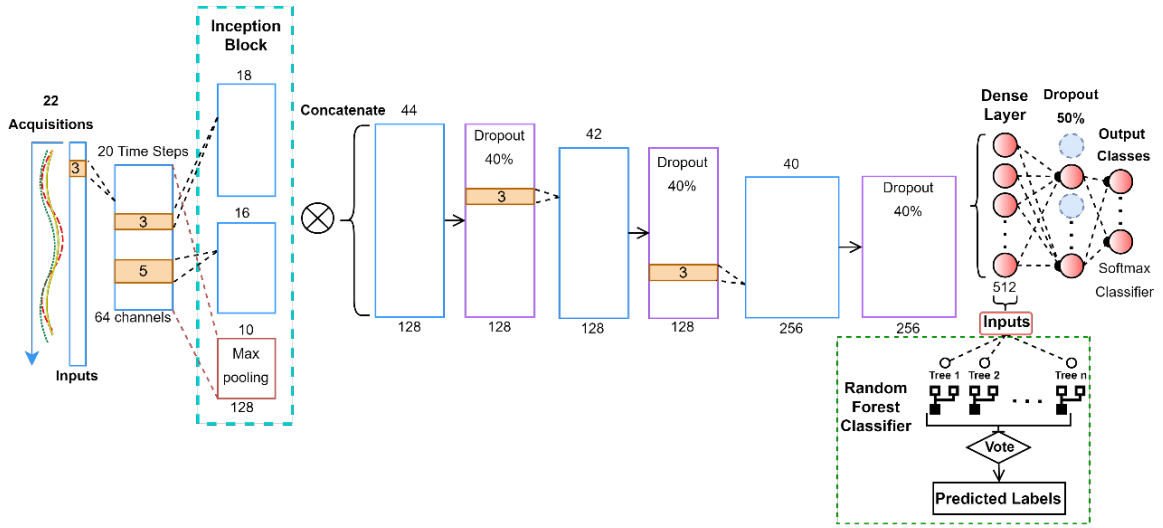

Figure 3-6. The architecture of Conv1D and Conv1D-RF. Conv1D processes original time-series inputs and determines output classes with the SoftMax classifier. Conv1D-RF extracts inputs from the dense layer and predicts using the random forest classifier.

3.2.6.3 Attention-based LSTM

The LSTM model, an extension of the conventional RNN, incorporates the concept of gates to synergistically process time-series data with long-term dependencies, retaining temporal features from the input sequence. Temporal features are stored in the LSTM cell at the current time step and passed to the next cell according to weights assigned by each gate, enabling the realization of sequential dependencies. This gate-to-gate delivery mechanism maintains the temporal features with previous and current states to a certain extent in each LSTM cell, allowing for long-and short-term memory formation (Zhou et al., 2016). Equations for LSTM gates are shown in Eq. (3-13) – (3-15).

$$f_t^c = \sigma(W_f \cdot [h_{t-1}^c, x_t^c] + b_f) \quad (3\text{-}13)$$

$$i_t^c = \sigma(W_i \cdot [h_{t-1}^c, x_t^c] + b_i) \quad (3\text{-}14)$$

$$o_t^c = \sigma(W_o \cdot [h_{t-1}^c, x_t^c] + b_o) \quad (3\text{-}15)$$

$$S_t^c = f_t^c \odot S_{t-1}^c + i_t^c \odot S_t'^c \quad (3\text{-}16)$$

$$S_t'^c = \tanh(W_S \cdot [h_{t-1}^c, x_t^c] + b_S) \quad (3\text{-}17)$$

$$h_t^c = o_t^c \odot \tanh(S_t^c) \quad (3\text{-}18)$$

Each LSTM cell unit has three gates, including forget gate f_t^c , input gate i_t^c, and output gate o_t^c. These gates use learnable weight matrices W_f, W_i, W_o and W_S with their biases b_f, b_i, b_o and b_S, to quantify information at the previous time step $t-1$ and the current time step t. The

information from the previous step is retained in f_t^c, updated in i_t^c, and then modulated by o_t^c before being passed to the next LSTM unit. The cell states S_t^c and the hidden state vectors h_t^c are updated through the current memory states $S_t'^c$ (Eq. (3-16) – Eq. (3-18)). C means the total number of feature bands. The symbol '·' is the product between matrices, while ⊙ denotes element-wise multiplication. The activation function σ is sigmoid used for gates, and tanh function updates the cell state.

The attention block refines aggregated hidden features learned by LSTM layers across a long time-series sequence for crop mapping by normalizing weights (Xu et al., 2020; Xu et al., 2021). Attention weights α_t^c are calculated using the Softmax activation function for normalising weight matrices W_α and bias b_α in Eq. (3-19) and Eq. (3-20). Hidden state vectors are updated as context vectors H_t^c, obtaining the final attention vectors using weight matrices W_H (Bahdanau et al, 2014) in Eq. (3-21):

$$\alpha_t^c = softmax(W_\alpha \cdot h_t^c + b_\alpha) \quad (3\text{-}19)$$

$$H_t^c = \alpha_t^c \cdot h_t^c \quad (3\text{-}20)$$

$$\boldsymbol{A}_t^c = softmax(W_H \cdot [H_t^c, h_t^c]) \quad (3\text{-}21)$$

3.2.6.4 Transformer

The transformer is a recent deep learning model for processing sequential data in natural language processing (NLP) (Vaswani et al., 2017). It also has been used for multi-temporal crop classification with optical data in agricultural remote sensing (Rußwurm and Körner., 2020; Xu et al., 2020; Xu et al., 2021). Figure 3-7 illustrates the custom Transformer network used in this study, consisting of positional embedding, encoding blocks, and a multi-layer perceptron (MLP) unit. Unlike NLP tasks, processing time-series remote sensing data does not require text embedding due to differences in data format and numerical significance. Instead, a one-dimensional position embedding is implemented for encoding and correlating time steps in time-series satellite data. Positional embeddings are linearly transformed into vectors for each attention head for self-attention computation. The self-attention mechanism focuses on interacting with positionally embedded vectors in a single sequence to compute the attended representation over the same sequence without requiring LSTM cell units to update hidden states.

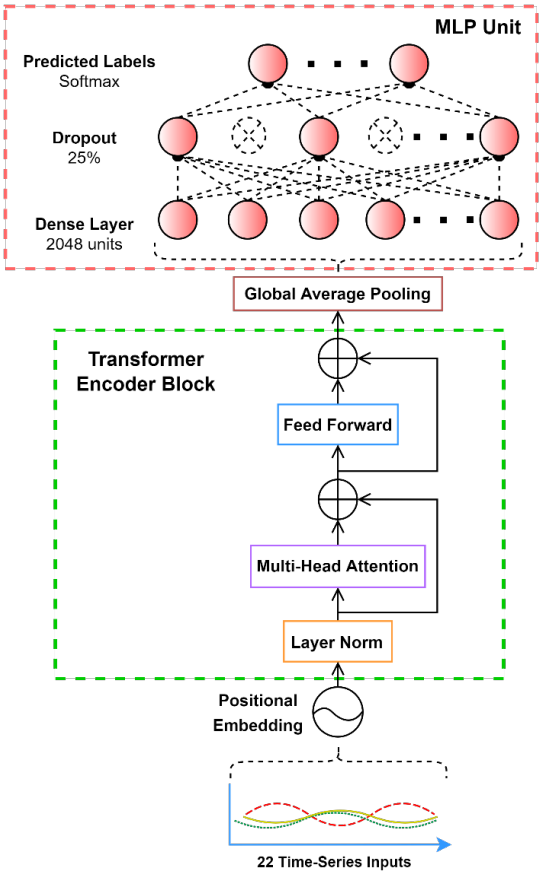

Figure 3-7. The transformer-based network in this study. It begins with positional embedding for encoding time-series inputs, progresses through the transformer encoder layers, and concludes with a multi-layer perception (MLP) for multi-class prediction.

The multi-head self-attention mechanism, consisting of Queries, Keys, and Values, captures different relationships or patterns within the input data. Employing multiple attention heads simultaneously allows the model to capture various aspects of the input data. However, self-attention may increase the volume of attention scores along the sequence length, resulting in a memory burden on computing over longer sequences (Xu et al., 2020). To optimize the network's performance and convergence, residual skip connections and normalization layers are applied before and after the multi-head attention layer. The output of multi-head attention is then transferred to a Feedforward module, which consists of linear layers with non-linear activation functions. This structure helps maintain the position-wise parallelism in the model and allows it to process long sequences more efficiently. Finally, the outputs of the encoder block are globally averaged over the temporal dimension and fed into the dense layers for the final prediction. Each encoder block has a head size of 256 and 8 attention heads. The total number of transformer encoder blocks is set to 2, followed by an MLP unit.

3.2.6.5 Conv1D-LSTM

In this study, we proposed and developed a joint ensemble learning strategy for improved crop classification performance, inspired by Zhou (2012) and its successful application in human activity recognition (Hamad et al., 2020). This method combines Conv1D and LSTM as parallel base learners for temporal feature extraction using customised internal components (Figure 3-8). This joint combination facilitates a mutually complementary effect by fusing multi-source features given by each module, thereby constructing a robust ensemble feature extractor for sequential data and mitigating model bias towards the majority class.

Firstly, we adapt the inception module design for processing multi-scale features (Zhong et al., 2019; Yang et al., 2020) within the 1D CNN module. We implement three Conv1D layers, each with kernel sizes of 3, 3 and 5 respectively with respect to incremental channel depths 64, 128 and 256 for feature learning at different scales, and use the skip connection technique at each level to alleviate the vanishing gradient problem (He et al., 2016). Batch normalisation (BN) is applied at the end of each Conv1D layer to accelerate model convergence. The output features from each level are fed into a fully connected layer and concatenated together.

Secondly, the LSTM module employs an attention-based mechanism with bidirectional LSTM cells for comprehensive temporal information capture (Yuan et al., 2020). The attention block aggregates sequential outputs from bidirectional LSTM, producing normalised attention weights to improve classification performance. A fully connected layer adjusts the outputs to 512 units, equal to the last output from the 1D CNN module. Finally, the learned features from each module are concatenated and passed through a shared fully connected layer, creating a strong ensemble feature extractor. A dropout layer prevents overfitting. and the final output layer generates class probabilities. The joint learning optimization, Conv1D-LSTM, maximizes parallel processing utilization for enhanced time-series crop type prediction.

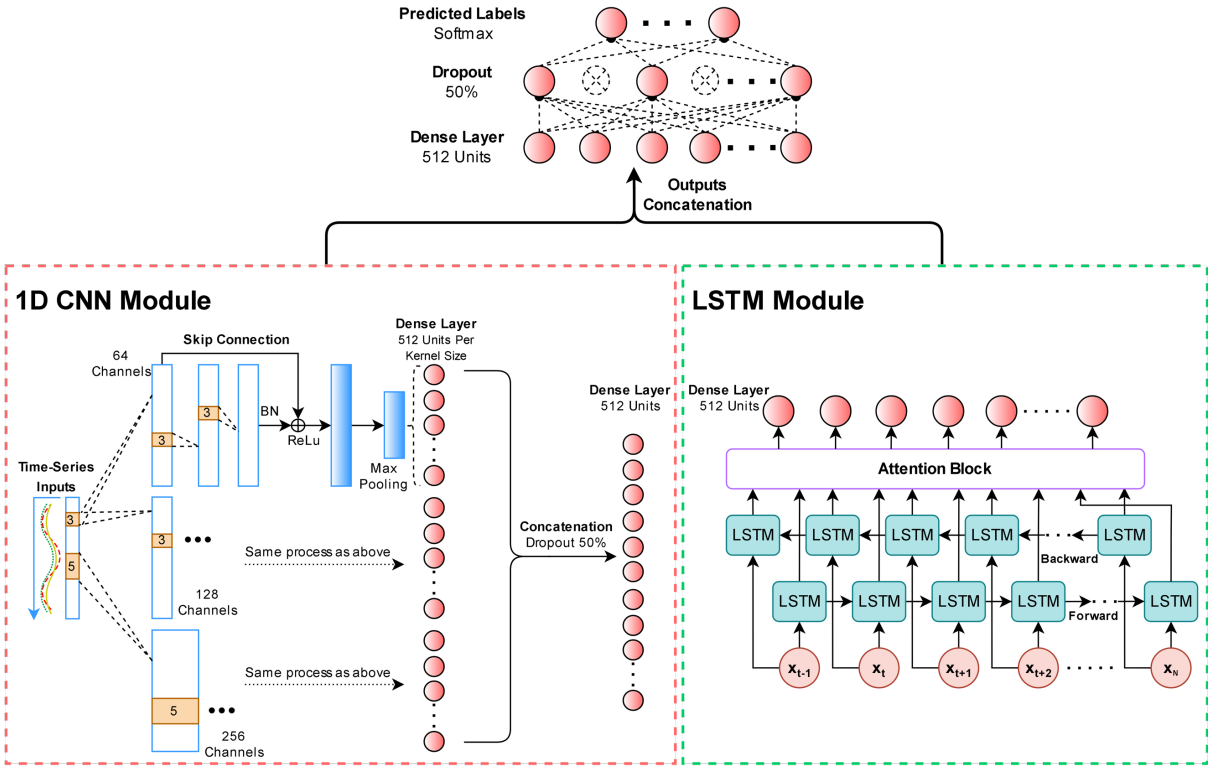

Figure 3-8. The proposed joint learning network: Conv1D-LSTM.

3.2.7 Model implementation

Three performance indicators were chosen to evaluate crop mapping performance: overall accuracy (OA), Cohen's kappa coefficient (Kappa), and F1 score. OA assesses the model's performance for all classes overall, as it represents the ratio of correctly classified samples to the total number of samples. Kappa adjusts classification scores by accounting for the probability of random chance prediction based on observed class frequencies in the dataset. The F1 score, which is the harmonic mean of recall and precision, measures classification performance for individual crop types. The deep learning model configurations are modulated with reference to the Adam optimizer (Kingma and Ba, 2014) and the cross-entropy loss function (Botev et al., 2013). The optimizer's learning rate is set at 0.001. The Rectified Linear Unit (ReLU) activation function (Nair and Hinton, 2010) is employed in all fully connected layers to enhance output nonlinearity and manage systematic errors (bias). Additionally, early stopping serves as a regularization strategy, halting model training when performance stabilizes during the iterative training process. All deep learning models aforementioned shared these configurations during the training stage. The modelling environment is implemented using Python 3.7.15 and Tensorflow 2.9.2, along with the Keras library (version 2.12) on a Windows

10 system equipped with two NVIDIA Quadro P4000 graphic devices (8 GB RAM per GPU) and two Intel (R) Xeon (R) Silver 4114 CPU processors (2.20GHz/2.19 GHz).

3.3 Results

3.3.1 Temporal profiles of Sentinel-1 SAR features

The temporal profiles of Sentinel-1-derived SAR features for each class, as displayed in Figure S1 in the Supplementary material, show the degree of separation between the crops, which appears challenging, as evidenced by the overlapping areas in the buffer zones across the full growing season for all SAR features. The similarity in the time-series profiles of the crops highlights the difficulty in differentiating between crop types with similar phenological cycles based on SAR-derived feature values. Although the values generally follow the growth trends towards their peaks around August, the separability between crops is not distinctly evident in specific SAR-based features. Some features, such as 'VH GLCM Mean', 'VH GLCM Variance', 'Entropy', and 'Alpha', exhibit fewer overlaps between August and September, during which maize, soybean, and other crops partially diverge as they approach the end of their growth stages. However, it remains uncertain whether these features alone can provide critical information for crop mapping during this period. The temporal profile of each SAR feature suggests a similar linear relationship between the features and corresponding crop types. Consequently, assessing the nonlinear relationship between variables and targets, as well as the correlation between SAR features, becomes necessary. Additionally, combining all of them increases feature dimensionality and computational cost. To address this, feature selection techniques were employed to filter out irrelevant features.

3.3.2 Feature selection outcomes

Boruta identifies optimal SAR features highly relevant to the prediction of target labels. As represented by grey circles (See Figure S2), those SAR features with the highest importance scores (normalised to 1) were identified by the Boruta wrapper model at corresponding points in time. Based on grey circles, we, as the user, manually selected SAR features falling within the entire crop growth cycle, which are crucial for in-season crop classification. These selected features are denoted by stars. Finally, the remaining features with complete temporal dimensions, such as DpRVI, GLCM, m-chi decomposition, and H/α dual-pol decomposition variables, were used to assess the multicollinearity among them.

The Spearman correlation coefficient (r_s) were calculated for each pair of selected features. This resulted in six pairwise combinations used to evaluate feature collinearity, characterized by each subplot (see Figure 3-9). Each axis limit represents the total number of acquisitions for a pair of combined features, with emphasis placed on evaluating correlation strength along the second interval on the x-axis and the first interval on the y-axis in each subplot, and vice versa. Each interval contains subclasses of a full-time series SAR feature. For instance, in the 'DpRVI with m-chi' subplot, the total number of features across the growth stage is 88, displayed on both the x-axis and y-axis. The first interval on each axis represents 22 columns of sequential data for DpRVI, while the second interval indicates 44 m-chi decomposition features in total. This arrangement forms a rectangle on each axis. All subclasses of each feature value are cross-calculated to determine the correlation strength, with the resulting r_s values derived from both features at the same time point represented by the off-diagonal lines of the rectangles. These lines display the intensity values relevant to the evaluation of correlation for the pairs of SAR features. In this case, the intensity values between DpRVI and m-chi in the same date range fall between 0.4 and 0.6, indicating a moderate level of correlation.

The ‘DpRVI with Entropy & Alpha’ pair exhibits stronger collinearity among all groups, as indicated by r_s intensity values exceeding 0.8 along the off-diagonal line of the rectangle and p-value less than 0.05. This suggests a high degree of correlation or linear dependency between the input variables used for classification, i.e., some of the features are redundant, as they provide similar or overlapping information about target labels. Although deep learning models and RF are less sensitive to collinearity, this study employed three scenarios for model accuracy assessment to mitigate its potential impact, which includes scenarios: (a) ‘GLCM’, ‘m-chi’, and ‘Dual-pol’; (b) ‘GLCM’, ‘m-chi’, and ‘DpRVI’; and (c) the combination of all features. Additionally, the individual feature was also evaluated for comparison.

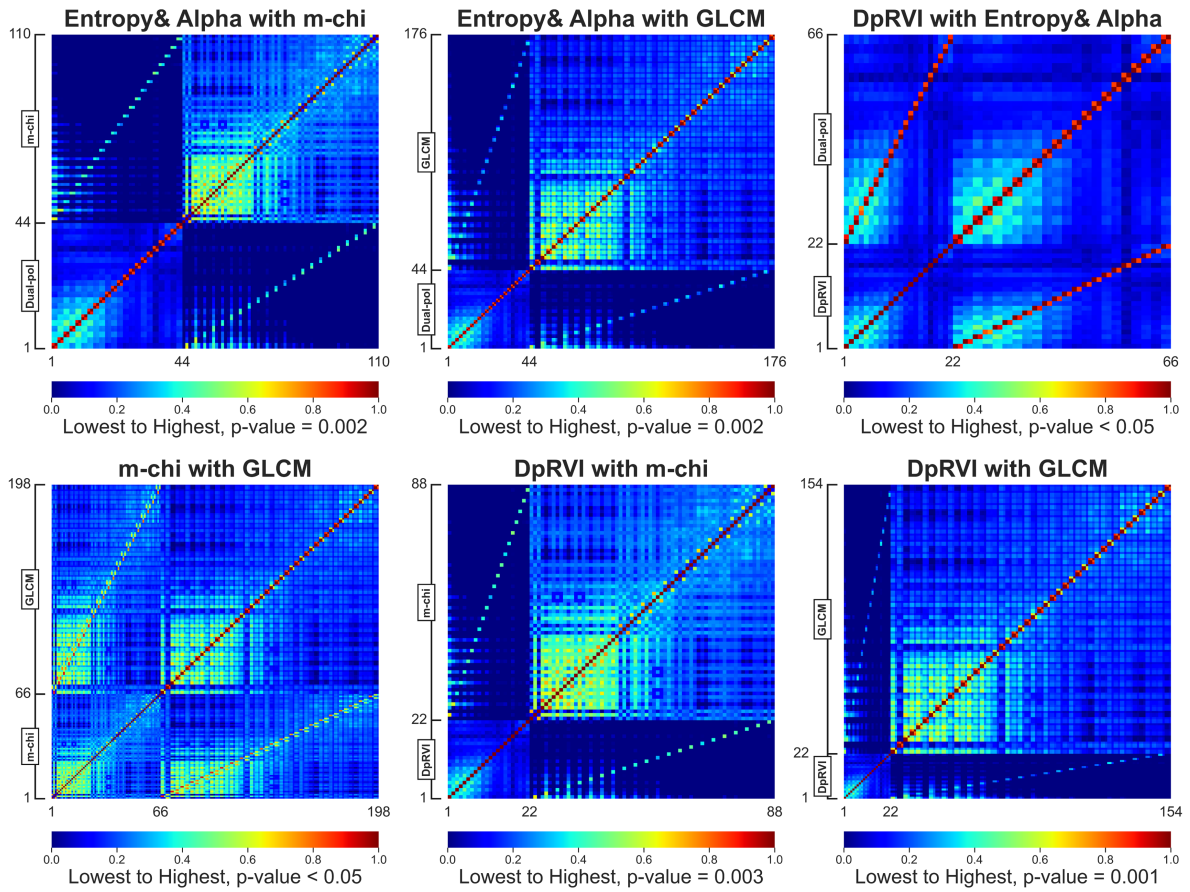

Figure 3-9. The adjacency matrix representation of Spearman rank correlation coefficient for the evaluation of the intervariable relationship. In each subplot, both axes represent the total number of pairwise combined features determined by the Boruta.

3.3.3 Accuracy assessment

The model performances are presented in Table 3-1, revealing that deep learning models generally outperform RF in terms of OA and Kappa, as demonstrated by Conv1D, Conv1D-RF, Transformer, and the proposed Conv1D-LSTM. Specifically, Conv1D-LSTM achieves the highest OA (0.88) using only m-chi features, while Conv1D, Conv1D-RF, and Transformer yield similar results but with the combined four features in scenario (c). It also results in the second-best Kappa (0.83), close to the best (0.84) led by scenario (c). This indicates that the diversity of input features may enhance accuracy for certain deep learning networks. However, models trained with optimal features can reduce computational costs for large-scale crop mapping. By selecting the most relevant and informative features, the complexity of the model is reduced, thus decreasing the time and resources required for training and prediction. This approach is particularly important when applying models to extensive datasets or in situations

where computational resources are limited. Consequently, scenario (c) exhibits redundancy in input predictors compared to using the m-chi variables alone.

Table 3-1. Accuracy assessment for selected features on testing sets. The model's training, validation, and testing were conducted using 10% of the entire sample data. The highest OA and Kappa values are highlighted in bold, and the second-best values are underlined.

OA					
Features	Conv1D	Conv1D-RF	Conv1D-LSTM	RF	Transformer
H/α	0.74	0.74	0.74	0.72	0.74
DpRVI	0.68	0.72	0.72	0.71	0.73
GLCM	0.80	0.82	0.82	0.79	0.83
m-chi	0.85	0.85	**0.88**	0.84	0.86
Scenario (a)	0.86	0.86	0.85	0.83	<u>0.87</u>
Scenario (b)	0.86	0.86	0.84	0.83	0.86
Scenario (c)	**0.88**	**0.88**	0.85	0.84	**0.88**
Kappa					
Features	Conv1D	Conv1D-RF	Conv1D-LSTM	RF	Transformer
H/α	0.64	0.64	0.64	0.62	0.65
DpRVI	0.58	0.61	0.75	0.60	0.62
GLCM	0.73	0.75	0.75	0.71	0.77
m-chi	0.79	0.80	<u>0.83</u>	0.78	0.80
Scenario (a)	0.81	0.81	0.79	0.77	0.82
Scenario (b)	0.81	0.81	0.78	0.77	0.81
Scenario (c)	<u>0.83</u>	<u>0.83</u>	0.80	0.78	**0.84**

Table 3-2 presents a comparison of model performance based on the F1 score for each class. Among the individual features, the m-chi feature demonstrates superior effectiveness for crop classification, as evidenced by higher F1 scores for models utilizing this feature compared to those using H/α, DpRVI, or GLCM. The proposed Conv1D-LSTM, when employing the m-chi features, exhibits the best performance, with the highest F1 scores for maize (0.85), soybean (0.84), other crops (0.82) and the highest average crop F1 score of 0.84. In scenarios (a), (b), and (c), the models generally achieve improved performance when combining multiple features, suggesting that incorporating a diverse set of input features can enhance overall F1 scores for certain deep learning networks. Wherein, scenario (c) is predominantly favoured by Conv1D, Conv1D-RF, Transformer, and RF, exhibiting the highest F1 scores compared to other scenarios for identifying crops and land cover types.

The Conv1D-LSTM using m-chi features highlights the importance of selecting optimal features to optimize computational efficiency in crop mapping and minimise redundancy in input predictors. Additionally, the computational cost for model training with m-chi features is substantially lower than that with scenario (c), as illustrated in Figure S3. Therefore, all ground-truth samples of m-chi features were trained, validated, and tested, following a 60%, 20%, and 20% ratio. The Conv1D-LSTM still yields the highest F1 scores for maize (87%), soybean (86%), and other crops (85%), outperforming other models (Figure 3-10). Overall, multi-stream deep learning architectures, such as the proposed Conv1D-LSTM and Conv1D-RF, exhibit superior performance compared to other standalone architectures, including Conv1D and Transformer. All deep learning networks consistently outperform the traditional RF in terms of the same metric for crop mapping.

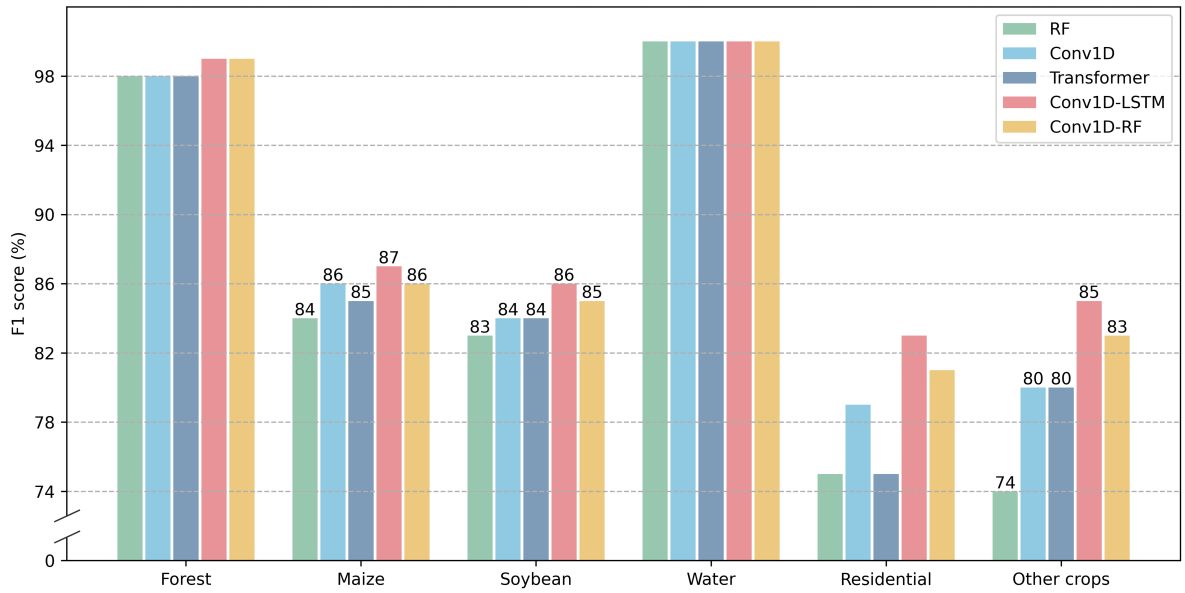

Figure 3-10. Comparison of model performance based on m-chi features. Models were trained with 60% of all ground-truth samples and tested on the same testing set. Values are displayed for crops.

Table 3-2. Model performance based on F1 scores. Highlighted in bold are the highest F1 scores for crop types, and the second-best scores are underlined. $\overline{Crops}$ denotes the average F1 scores for all crop classes.

Model	Features	Residential	Forest	Maize	Soybean	Water	Other crops	$\overline{\boldsymbol{Crops}}$
Conv1D (Zhong et al., 2019)	H/α	0.74	0.84	0.71	0.71	0.98	0.45	0.62
	DpRVI	0.29	0.80	0.59	0.69	0.92	0.35	0.54
	GLCM	0.58	0.97	0.73	0.77	1.00	0.67	0.72
	m-chi	0.70	0.98	0.82	0.80	0.99	0.70	0.77
	Scenario (a)	0.77	0.98	<u>0.84</u>	0.82	1.00	0.75	0.80
	Scenario (b)	0.75	0.98	0.82	0.82	1.00	0.75	0.80
	Scenario (c)	0.81	0.99	**0.85**	**0.84**	1.00	0.79	<u>0.83</u>
Conv1D-RF (Yang et al, 2020)	H/α	0.61	0.84	0.71	0.72	0.98	0.48	0.64
	DpRVI	0.37	0.82	0.70	0.70	0.97	0.43	0.61
	GLCM	0.64	0.97	0.78	0.77	1.00	0.71	0.75
	m-chi	0.75	0.98	0.82	0.81	1.00	0.74	0.79
	Scenario (a)	0.76	0.98	0.83	0.82	1.00	0.75	0.80
	Scenario (b)	0.75	0.98	0.83	0.82	1.00	0.75	0.80
	Scenario (c)	0.80	0.99	**0.85**	**0.84**	1.00	0.79	<u>0.83</u>
RF	H/α	0.26	0.83	0.70	0.70	0.96	0.37	0.59
	DpRVI	0.15	0.81	0.69	0.69	0.96	0.37	0.58
	GLCM	0.34	0.96	0.75	0.74	1.00	0.58	0.69
	m-chi	0.62	0.97	0.81	0.80	1.00	0.64	0.75
	Scenario (a)	0.56	0.97	0.81	0.80	1.00	0.59	0.73
	Scenario (b)	0.54	0.97	0.81	0.79	1.00	0.60	0.73
	Scenario (c)	0.66	0.97	0.82	0.81	1.00	0.66	0.76
Conv1D-LSTM	H/α	0.65	0.84	0.71	0.72	0.98	0.44	0.62
	DpRVI	0.36	0.82	0.70	0.70	0.97	0.43	0.61
	GLCM	0.58	0.97	0.78	0.76	1.00	0.69	0.74
	m-chi	0.79	0.98	**0.85**	**0.84**	1.00	**0.82**	**0.84**
	Scenario (a)	0.73	0.98	0.83	0.81	0.99	0.63	0.76
	Scenario (b)	0.63	0.97	0.81	0.81	1.00	0.67	0.76
	Scenario (c)	0.74	0.98	0.83	0.81	1.00	0.70	0.78
Transformer (Rußwurm and Körner, 2020)	H/α	0.65	0.84	0.72	0.72	0.98	0.48	0.64
	DpRVI	0.45	0.82	0.70	0.70	0.97	0.45	0.62
	GLCM	0.66	0.98	0.80	0.78	1.00	0.73	0.77
	m-chi	0.67	0.97	0.83	0.82	1.00	0.72	0.79
	Scenario (a)	0.77	0.98	<u>0.84</u>	<u>0.83</u>	1.00	0.76	0.81
	Scenario (b)	0.75	0.98	0.83	0.82	1.00	0.76	0.80
	Scenario (c)	0.75	0.99	**0.85**	**0.84**	1.00	<u>0.80</u>	<u>0.83</u>

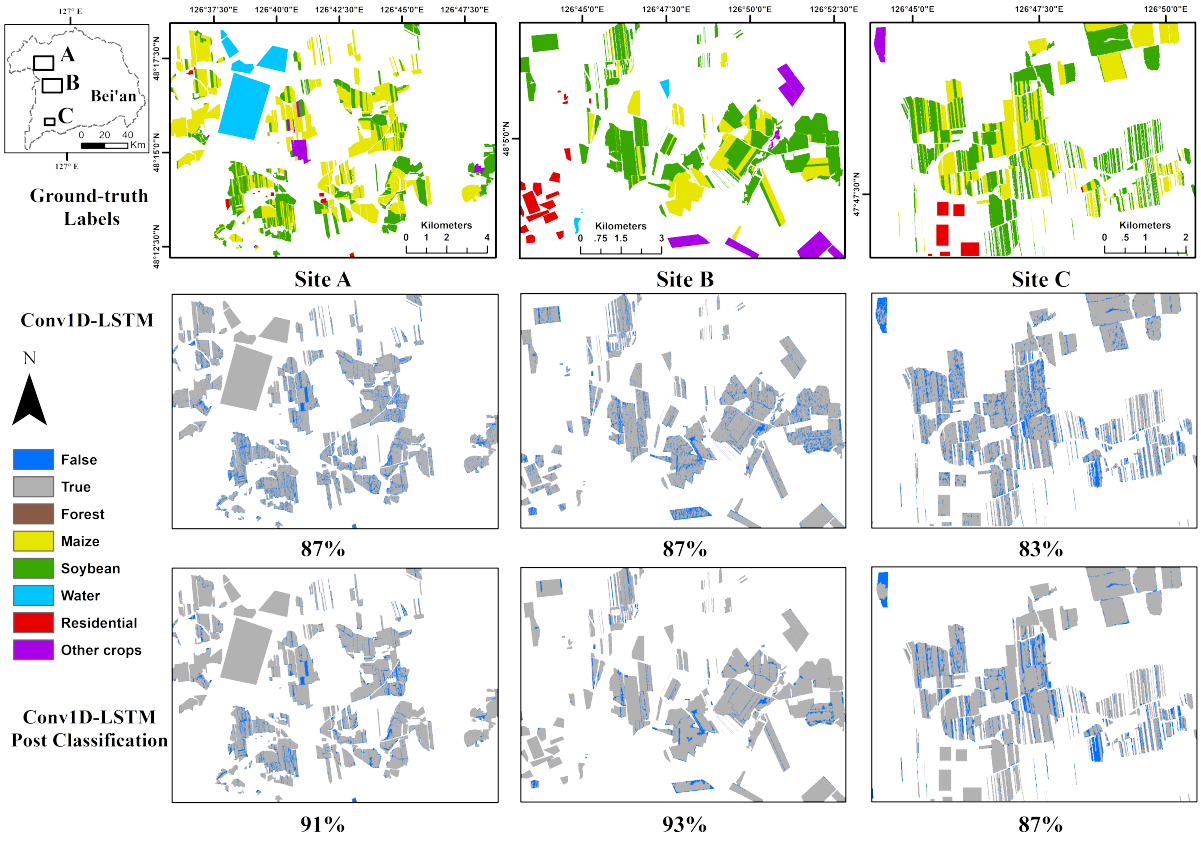

Figure 3-11. Comparison of Conv1D-LSTM performance in three sample sites before and after post-classification. Percentages indicate the ratio of correctly classified pixels to ground-truth labels.

Given that the optimal SAR feature is decided, the model generalisability was assessed across multiple small-scale regions within Bei'an. With 60% of all samples randomly selected from m-chi decomposition features for model training, Conv1D, RF, transformer, and Conv1D-LSTM were employed for spatial map predictions and classification comparisons. The qualitative results for different geospatial locations are shown in Figure S4, Figure S5 and Figure S6.

Further post-classification procedures based on majority selection were conducted to reduce misclassified pixels in the prediction maps derived by Conv1D-LSTM (See Figure 3-11), which demonstrates improved classification performance across all three sites compared to the proposed method alone. As a result, a county-scale classification map was produced by synergistically utilising the Conv1D-LSTM pre-trained based on 60% of all samples and the post-classification technique, as shown in Figure 3-12. The improvements in classification performance can be observed in the confusion matrices (See Figure 3-13). Specifically, the recall values of maize, soybean and other crops increased by 4%, 8% and 7%, respectively in

the post-classification process, which suggests the effectiveness of combining Conv1D-LSTM with post-classification techniques in improving the overall classification accuracy.

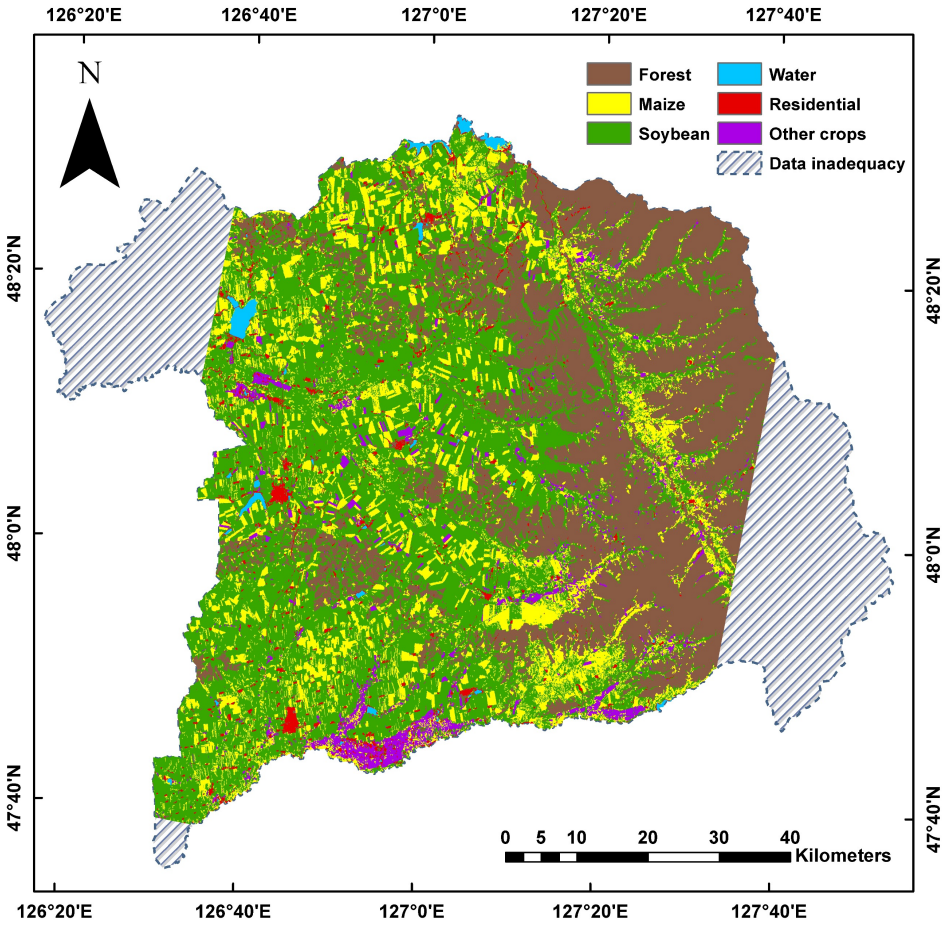

Figure 3-12. Post-classification prediction map for Bei'an 2017. Data inadequacy refers to missing data resulting from incomplete coverage of the whole study area in Sentinel-1 SLC acquisitions across the growth season.

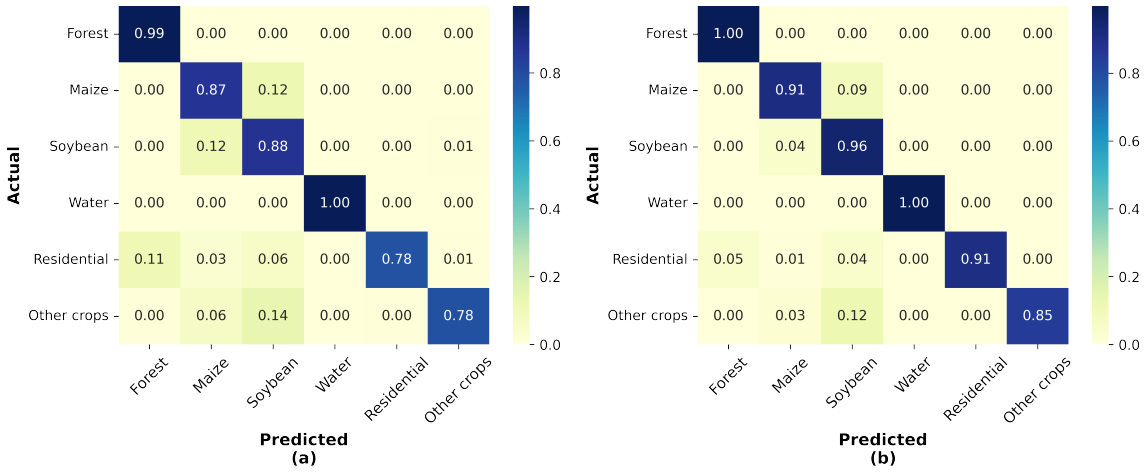

Figure 3-13. Normalised confusion matrix for Conv1D-LSTM. (a) pre-processed, (b) post-processed. Values are normalized as percentages for each class.

3.3.4 In-season crop mapping

In-season crop mapping was assessed based on monthly-based temporal blocks corresponding to the phenology of major crops. This approach simulates the practical scenario, considering the increased availability of satellite acquisitions throughout the growing season. As Conv1D-LSTM outperformed other Conv1D-based architectures according to section 2.3, the machine learning model RF and the Transformer, which relies on the self-attention mechanism, were used for comparison with Conv1D-LSTM in the assessment of in-season crop mapping (Table 3-3). The Conv1D-LSTM model surpassed the RF and Transformer models in terms of overall accuracy (OA) across all temporal windows. The OA for Conv1D-LSTM increased from 0.69 in May to 0.88 in September, emphasizing the significance of the temporal dimension for sequential networks. For all models, the F1 scores for maize and soybean generally increased over the temporal windows, with the Conv1D-LSTM model exhibiting the highest F1 scores for both crop types in each month.

Improvements for maize were observed during the earlier growth stages, with Conv1D-LSTM, RF, and Transformer improving by 8%, 9%, and 8% in June (tasselling), respectively, and by 9%, 6%, and 8% in July (ripening). As September (harvest) approached, moderate improvements of around 3% were recorded for all models in August (end of flowering or silking) and September. Similarly, early growth stage improvements for soybean were 9%, 7%, 7% and 5%, 4%, and 6% in June (flowering) and July (podding and seed filling), respectively, for each model. Values continued to improve by 5%, 4%, and 5% in August (matured), followed by slight improvements (3%, 3%, 2%) in September (harvest). The most notable increase in F1 scores for other crops was observed between May and June, with substantial improvements in June accounting for 42%, 33%, and 44% for Conv1D-LSTM, RF, and Transformer, respectively. August yielded the lowest improvement for other crops (Conv1D-LSTM: 5%, RF: -1%, Transformer: 5%), but the transition from August to September improved models by 11%, 19%, and 7%.

Table 3-3. In-season crop mapping performance. Model training, validation and testing were based on 10% of all sample data of m-chi decomposition features. Highlighted in bold are the best results.

Temporal windows			May	June	July	August	September
Conv1D-LSTM	OA		0.69	0.76	0.82	0.85	**0.88**
	F1 scores	Maize	0.61	0.69	0.78	0.82	**0.85**
		Soybean	0.62	0.71	0.76	0.81	**0.84**
		Other crops	0.12	0.54	0.66	0.71	**0.82**
RF	OA		0.68	0.75	0.79	0.82	0.84
	F1 scores	Maize	0.60	0.69	0.75	0.79	0.82
		Soybean	0.62	0.69	0.73	0.77	0.80
		Other crops	0.09	0.42	0.46	0.45	0.64
Transformer	OA		0.68	0.74	0.80	0.84	0.86
	F1 scores	Maize	0.59	0.67	0.75	0.81	0.83
		Soybean	0.62	0.69	0.75	0.80	0.82
		Other crops	0.03	0.47	0.60	0.65	0.72

3.4 Discussion

3.4.1 Impact of SAR and temporal features on model performance

In addressing the first research question, this study found m-chi decomposition features to be efficient and effective in distinguishing crops, in agreement with previous findings (e.g. De et al., 2014; Sonobe et al., 2019; Mahdianpari et al., 2019; Dingle et al., 2022). Sonobe et al (2019) reported an F1 score of 0.84 for maize using RF, which was marginally lower than the scores of 0.85 (Table 3-2) and 0.87 (Figure 3-10) obtained in this study for maize before post-classification. However, they combined multiple predictors, ranging from 8 to 11, including the RVI, backscatter coefficients, and m-chi features, while this study relied solely on m-chi features. Although incorporating additional features could improve performance in scenario (c) in Table 3-2, it could lead to extended model training time (Figure S3). Using only m-chi features, this study's performance is comparable to that of Mahdianpari et al (2019), who produced recall values for maize (0.92) and soybean (0.79) using object-based RF, compared to maize (0.91) and soybean (0.96) derived in this study (see Figure 3-13). In the current study, covariance matrix parameters and backscatter coefficients were not included in the assessment of model accuracy according to Boruta (Figure S2); however, their importance still supports earlier findings in crop classification (e.g., Qu et al., 2020; Sun et al., 2019; Moumni and Lahrouni, 2021).

Previous studies have utilised fully polarimetric data from coherency matrices (T_3) for multi-temporal crop classification based on deep learning approaches. For instance, Liao et al. (2020) achieved the best F1 scores of 0.93 and 0.92 for maize and soybean, respectively, using

decomposed SAR features derived from quad-pol data. The values surpass the best F1 scores for maize (0.87) and soybean (0.86) before post-classification in the current study. The potential explanation for this discrepancy may be that T_3 provides richer polarimetric information than C_2, increasing the likelihood of identifying complex land covers. Similarly, He et al. (2020) and Xie et al. (2019) also extracted polarimetric features from T_3 using various quad-pol decomposition algorithms for crop mapping and achieved accuracies above 0.90 for both maize and soybean. However, He et al. (2020) applied the transfer learning techniques to learn scarce PolSAR features for crop classification. Xie et al. (2019) extracted data from RADARSAT-2 (fine-quad wide beam mode) with a higher spatial resolution (5 m) than Sentinel-1 (10 m) data applied in this study. Although exploiting quad-pol data could result in better performances than using dual-pol features for crop mapping, space-based quad-pol sensors are substantially limited by swath coverage (Raney, 2019), constraining their applicability for map prediction at the county level. Additionally, obtaining time-series quad-pol data might not be financially viable.

The models employed in this study for comparisons, such as Conv1D (Zhong et al., 2019), Conv1D-RF (Yang et al., 2020), and Transformer (Rußwurm and Körner, 2020), were initially designed for multispectral data and vegetation indices for crop mapping. Although these models have been successfully demonstrated in classifying crops using optical data, the primary focus of the present study is on SAR-related features. Nevertheless, the Conv1D-LSTM combined with optimal SAR features outperformed other model architectures, reflecting the findings of Hamad et al. (2020) regarding joint ensemble learning networks. These networks efficiently capitalize on the strengths of two sequential models, such as Conv1D and attention-based LSTM, to tackle the imbalanced class distribution frequently observed in real-world crop datasets, which contain both majority and minority crop classes. This corresponds to the second research question outlined in Section 1.

The in-season classification scheme (Section 3.3.4) operates model training based on incremental time-series data along growth stages. Model performances in earlier growth stages were not statistically significant, suggesting that the SAR features available from maize with leaves and collars, as well as emerging soybeans, were insufficient for effective model extraction. The poor performance of other crops in the early season can be attributed to the impact of bare soil. However, performance gradually improves as acquisitions corresponding to phenological stages are continuously accumulated. This supports the hypothesis that

consecutive temporal windows in line with complete growth stages are inherently beneficial for enhancing crop mapping. This finding is consistent with previous studies (Wei et al., 2019; Xu et al., 2020), which demonstrated that each growth stage of observations in a complete time series enriches data structures during model training, which corresponds to the third research question. Although all models benefited from sequentially accumulated SAR acquisitions, Conv1D-LSTM consistently outperformed RF and Transformer on a monthly basis in terms of recognizing all crop types.

3.4.2 Interpretation of learning behaviour of Conv1D-LSTM

The interpretability of deep learning models remains a challenge in remote sensing studies. In the end-to-end neural networks, weights are iteratively updated and optimized through backpropagation to minimize errors between inputs and targets. The full set of mathematical operations is concealed during model training, resulting in complexities and difficulties in outcome analysis, commonly known as the "black box." Although many studies have thoroughly explained Conv1D-based architectures and attention-based LSTM in terms of their learning behaviours with time-series multi-spectral data for crop classification (Zhong et al., 2019; Yang et al., 2020; Rußwurm and Körner, 2020; Xu et al., 2021), the proposed joint ensemble learning architecture in this study (Conv1D-LSTM), due to its intricate design, necessitates interpretation of the learned hidden features in the 1D-CNN and LSTM module to respond the fourth research question.

The visualization of the learning behaviour of the 1D-CNN end of Conv1D-LSTM is intuitively assessed in Figure 3-14. Weight distribution along time steps for each class is visualized over multi-level Conv1D layers. Selected convolutional layers include the 1st layer with 64 channels, the 2nd layer with 128 channels, and the 3rd layer with 256 channels, allowing for interpretation of the model on multi-scale feature learning. For crops, weights are mostly distributed around '08-15' in the 1st layer, suggesting that important acquisitions could be roughly identified by the model's shallow layer. As feed-forward propagations advance through the following layers, weight distributions become more localized, and important acquisitions are identified at specific time steps. For instance, more acquisitions are intensively weighted around June (tasselling for maize and flowering for soybean) and July (maturing) for crops. Comparing this finding with those of studies (Zhong et al., 2019; Yang et al., 2020) confirms that simpler patterns of features typically respond to shallow Conv1D layers and complex patterns are

presented by deep layers due to multi-scale feature learning. Another visualization of weight distribution is based on the average weight on each time step (see Figure 3-15). Weight intensities appear to fluctuate diffusely throughout the entire growth stage as the layer goes deep, with important acquisitions identifiable at the 3rd Conv1D layer for each class. However, regarding the trends for crops, their similar weight profiles may still contribute to high rates of model misclassification. This can be modulated with the LSTM module of Conv1D-LSTM.

The attention-based LSTM in Conv1D-LSTM generates attention weights that can be observed in Figure 3-16. A distinct difference in weight profiles for maize and soybean is evident, potentially enhancing the separability between these two crops. This outcome is comparable to a previous study by Xu et al (2021), which presented similar attention weight distributions for maize and soybean when using multi-spectral raw bands as inputs for the attention-based LSTM. This further emphasizes the significance of m-chi features in distinguishing crops based on the model's decisions. The attention weights for maize and other crops start to increase rapidly from '07-10' as they enter the maturation stage, implying that cumulative information becomes progressively more useful for crop classification from this point onwards. For soybean, the same turning point is observed around '08-15' during its maturity stage. These results align with the findings of Xu et al (2021), who also identified rapidly increasing attention weights for maize at the silking stage on '07-15' (following the tasselling stage) and for soybean on '08-19'. Moreover, it is plausible that the observed increase in the F1 score for other crops in September (Table 3-3) may be connected to the rising attention weights within the same temporal window since the weight intensities are not significant in the Conv1D layers around '09-15'. This observation could help elucidate the attention mechanism's contribution to classifying the minority class.

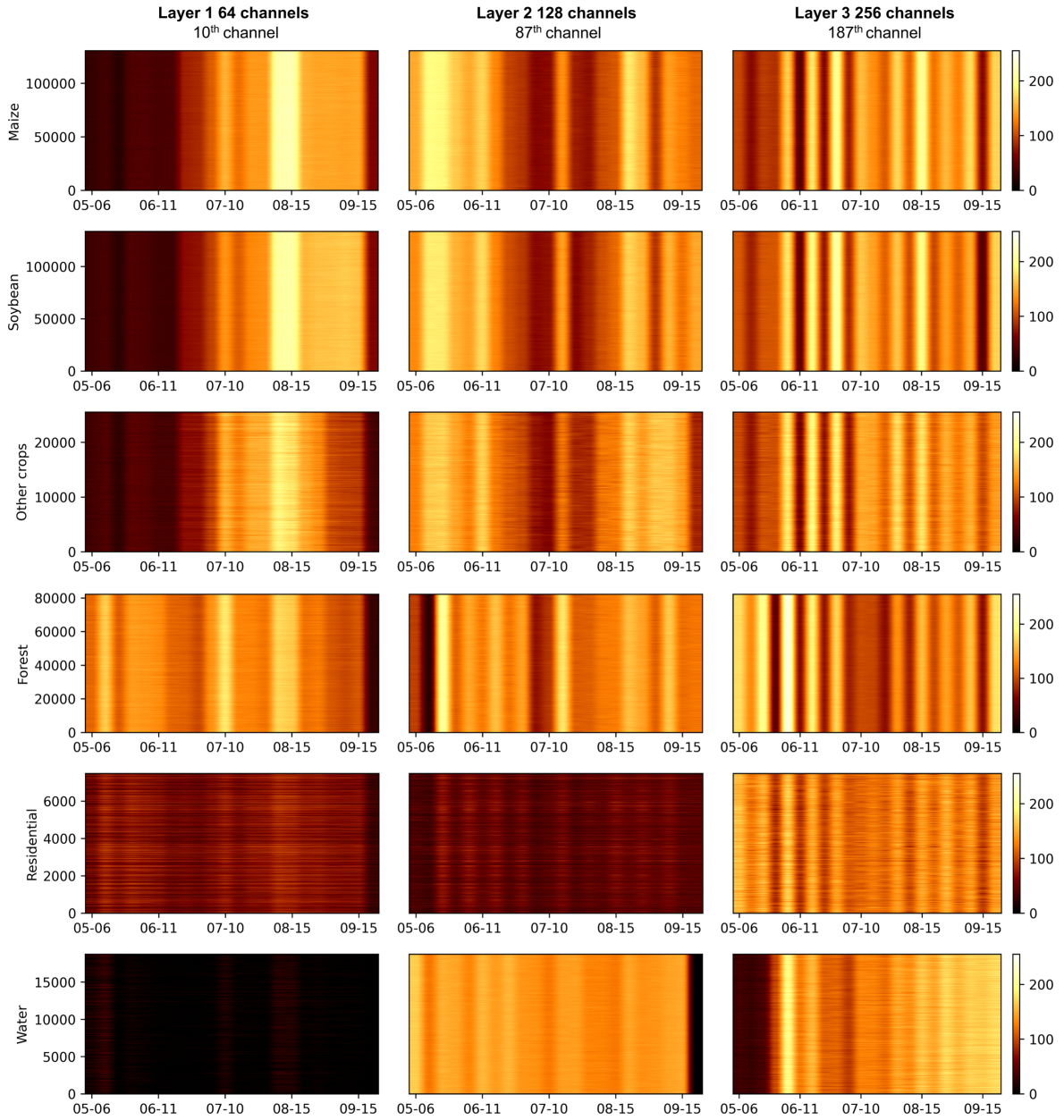

Figure 3-14. Visualisation of feature maps from three-level Conv1D layers in Conv1D-LSTM. The output feature maps were extracted from the model pre-trained with m-chi features. The y-axis represents the sample count for each class in the testing set, while the x-axis displays Sentinel-1 acquisitions at monthly intervals. The weight values' range is normalized from 0 to 255 to facilitate visualizing intensity for weight distributions within each channel.

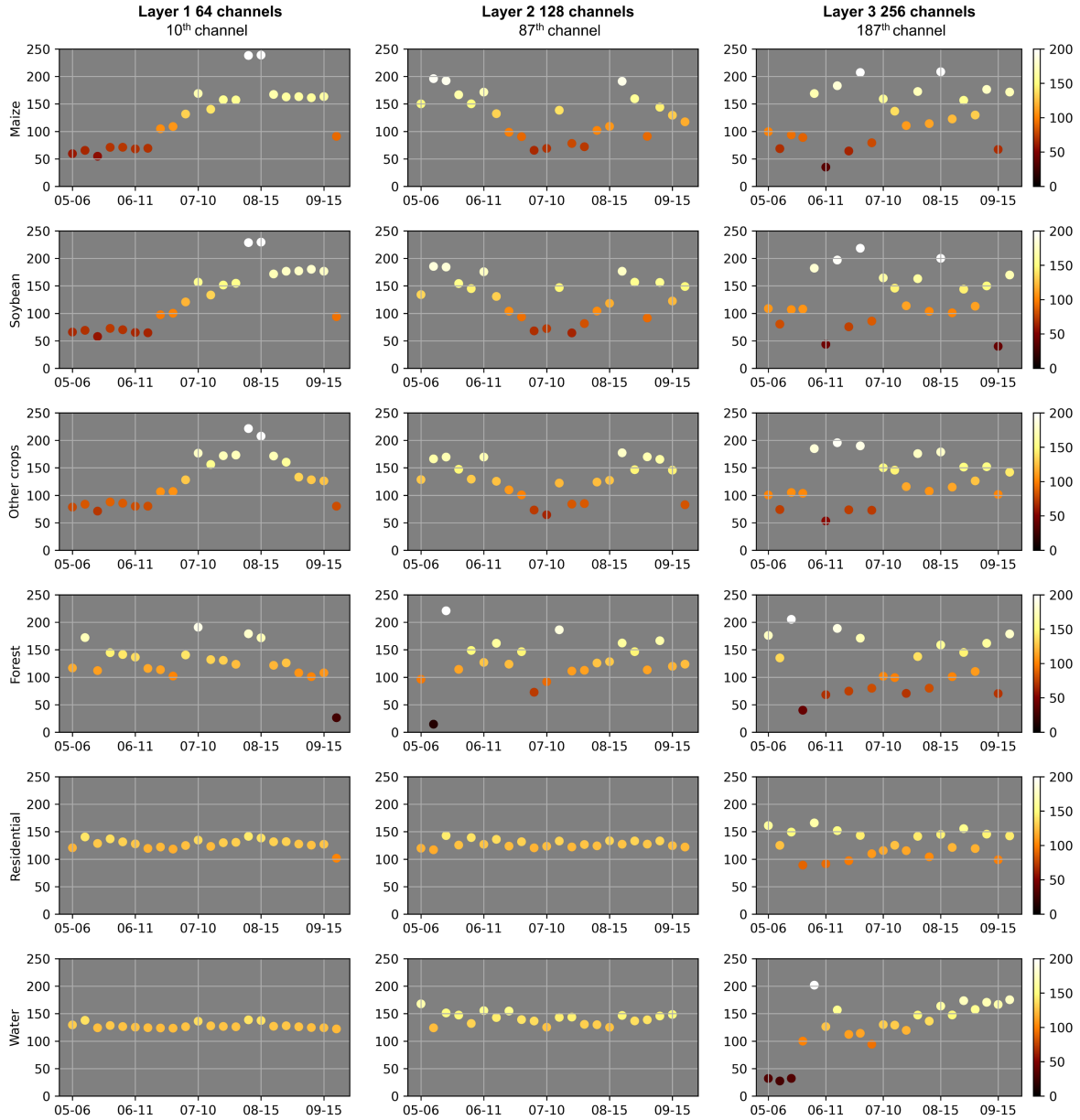

Figure 3-15. Visualisation of feature maps based on average weight distribution across multi-scale Conv1D layers. The range of weight values is normalized from 0 to 255 on the y-axis.

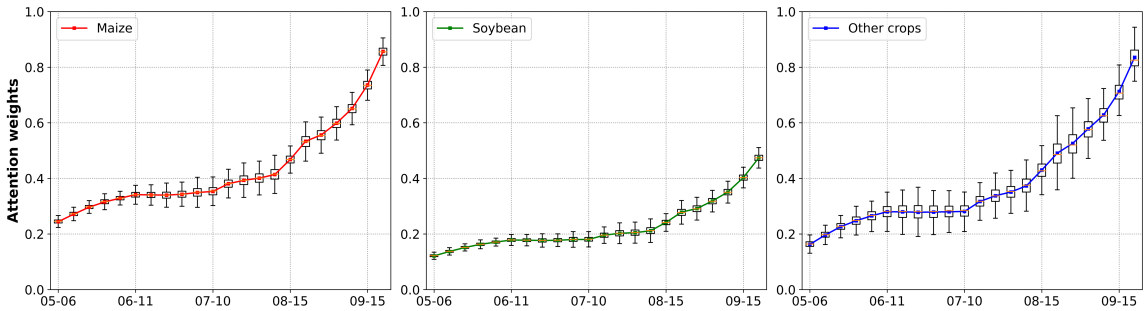

Figure 3-16. Crop attention weight profiles derived from the LSTM module in Conv1D-LSTM. The output attention weights were obtained using m-chi features. Values on the y-axis are scaled to a range from 0 to 1 for improved visualisation.

The model performance is not only reflected in the model accuracy assessment but can also be visualized through learned hidden features. The t-distributed stochastic neighbour embedding (t-SNE) visualization technique, introduced by Van der Maaten and Hinton (2008), allows for dimensionality reduction on high-dimensional features. In this study, the high-level hidden features are 512-dimensional and exhibit complex patterns for multi-temporal information. Analysing such a data structure intuitively is challenging, but t-SNE enables the nonlinear projection of hidden features onto a two-dimensional plane for visual comparison of crops' separability along monthly blocks (See Figure 3-17). In the early stage in May (emerging), the learned neural samples of all crops are closely situated. As the growth stages progress, these samples become increasingly segregated and grouped according to crop types. From June (tasselling and flowering) onward, crop separability expands toward the harvest stage in September, demonstrating that the accumulation of temporal information contributes to improving crop classification performance. These findings align with the in-season classification results in Table 3-3. Although t-SNE is not always a conclusive indicator for supporting quantitative analysis of learned features, it provides an alternative way to intuitively examine the outcomes of the hidden learning patterns from deep learning models.

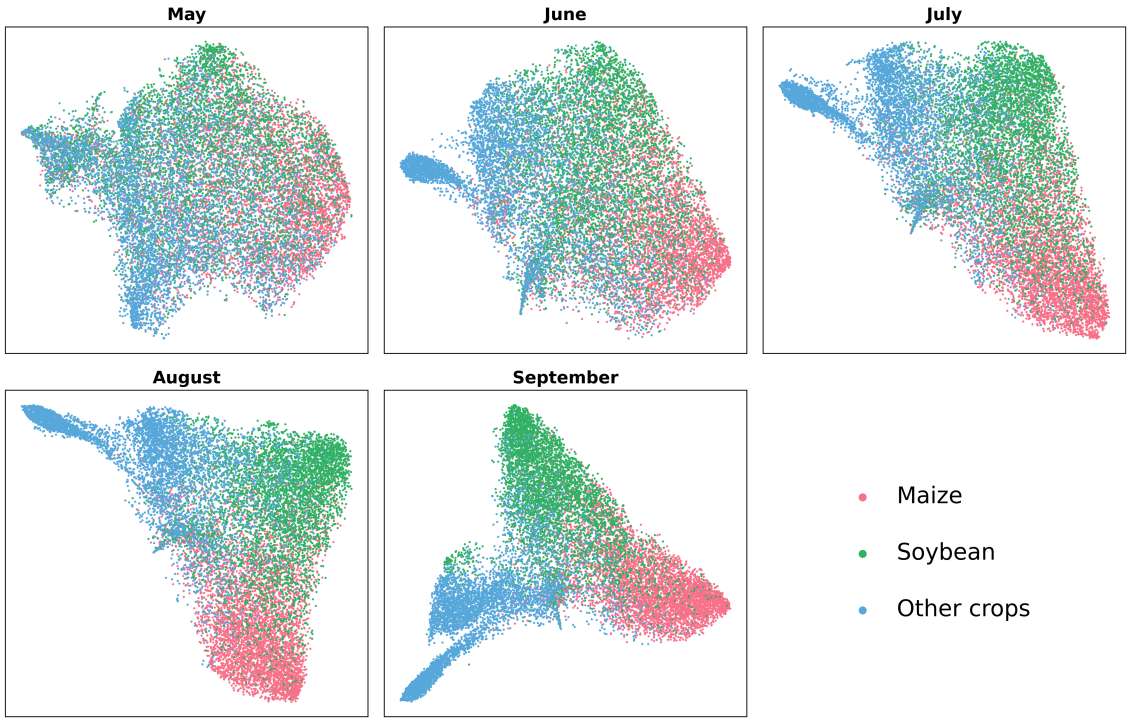

Figure 3-17. Visual comparison based on t-SNE along monthly blocks for learned hidden features. Wherein 5000 randomly selected neural samples from each crop type were extracted from the MLP unit of the Conv1D-LSTM model to enhance visualization. Each point represents a single neural sample corresponding to a specific crop category.

3.4.3 Potential extension of this study

In this study, the dual-polarisation system of Sentinel-1 SLC has demonstrated its potential to generate pseudo compact PolSAR parameters, which, to a certain extent, can improve the discrimination for crops with phenological similarities. This is achieved with the aid of the proposed joint ensemble learning Conv1D-LSTM for enhanced classification performance. This suggests that compact polarimetry holds promise for improving land cover classification. Moreover, there is a need for an expanded time series to correlate with more detailed phenological phases of crops. This could be accomplished by incorporating a few optical acquisitions that align with key SAR acquisitions identified in this study, which relates to the synergistic use of SAR and multi-spectral features for improved crop mapping (Liao et al., 2020).

Further optimization of the model architecture and classification workflows can enhance multi-temporal crop mapping. For instance, Conv1D-LSTM focuses merely on pixel-based feature extraction from time-series data, disregarding spatial relationships between features. However, 3D-CNN architectures, which have shown promise in crop classification using optical data (Ji et al., 2018), take both temporal and spatial dimensions into account. Exploring this approach for crop mapping that combines SAR and optical data, such as in the study by Teimouri et al (2022), could lead to improvements. Another promising avenue for model development is the integration of data-driven models and physical models for agricultural applications using remote sensing data. Physical models offer strong interpretability and performance but often contain redundant parameters and lack efficiency. In contrast, deep learning models suffer from limited interpretability. Combining both model mechanisms could create a mutually complementary effect, addressing each other's limitations. Existing studies have applied this combination by replacing a submodule of a physical model with RF (Keller and Evans, 2019) or by modulating the loss function in the deep learning model using physical mechanisms (Yang et al., 2023). It would be promising to examine how these approaches could potentially improve the performance of crop mapping. In addition to model design, it is also essential to evaluate the spatiotemporal transferability of the models in crop mapping, considering interregional and interannual variability.

3.5 Conclusion

This study has successfully demonstrated the potential of utilizing joint ensemble learning of two temporal models for extracting multi-temporal features from dual-pol SAR data to predict county-level crop categories for Bei'an in 2017. The proposed Conv1D-LSTM model has shown its efficiency and effectiveness, outperforming previously validated Conv1D, Conv1D-RF, Transformer, and baseline RF models. With optimally selected SAR features (m-chi decomposition), the Conv1D-LSTM achieved the highest F1 scores (87%, 86%, and 85%) for maize, soybean, and other crops, respectively. The results also highlight the model's ability to handle inherently imbalanced data and differentiate between summer crops with similar phenology. The importance of multi-temporal information for crop classification was emphasized by the in-season classification results. While all selected models benefited from increased acquisitions, the Conv1D-LSTM effectively captured temporal dependencies across complete growth stages, leading to superior monthly performance compared to other models. Post-classification further enhanced classification performance based on the proposed model. Furthermore, this study provided multiple perspectives on the model learning process by identifying critical phenological stages through visualisations of weight distributions at each end of the Conv1D-LSTM architecture. Hidden feature analysis unveiled the learning impact of the Conv1D-LSTM throughout monthly temporal intervals, indicating that temporal model performance in crop mapping depends on diverse phenological characteristics within time-series data. Ultimately, the joint learning of Conv1D and attention-based LSTM exemplifies the considerable potential to produce accurate cropland data layers at a large scale. In the subsequent stage, we aim to design a spatiotemporal learning approach, building on our current framework, to optimally transfer reliable pixels and pre-trained models from Bei'an to other research regions.

References

Ainsworth, T.L., Kelly, J.P. and Lee, J.S., 2009. Classification comparisons between dual-pol, compact polarimetric and quad-pol SAR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(5), pp.464-471.

Azzari, G., Jain, M. and Lobell, D.B., 2017. Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. *Remote Sensing of Environment*, *202*, pp.129-141.

Bahdanau, D., Cho, K. and Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Baloloy, A.B., Blanco, A.C., Candido, C.G., Argamosa, R.J.L., Dumalag, J.B.L.C., Dimapilis, L.L.C. and Paringit, E.C., 2018. Estimation of mangrove forest aboveground biomass using multispectral bands, vegetation indices and biophysical variables derived from optical satellite imageries: Rapideye, planetscope and sentinel-2. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*, pp.29-36.

Bargiel, D., 2017. A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sensing of Environment*, *198*, pp.369-383.

Belgiu, M. and Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, pp.24-31.

Benos, L., Tagarakis, A.C., Dolias, G., Berruto, R., Kateris, D. and Bochtis, D., 2021. Machine learning in agriculture: A comprehensive updated review. *Sensors*, *21*(11), p.3758.

Beriaux, E., Jago, A., Lucau-Danila, C., Planchon, V. and Defourny, P., 2021. Sentinel-1 time series for crop identification in the framework of the future CAP monitoring. *Remote Sensing*, *13*(14), p.2785.

Bertsimas, D., Pawlowski, C. and Zhuo, Y.D., 2018. From predictive methods to missing data imputation: an optimization approach. *Journal of Machine Learning Research*, *18*(196), pp.1-39.

Bexell, M. and Jönsson, K., 2017, January. Responsibility and the United Nations' sustainable development goals. In *Forum for Development Studies*, *44*(1), pp. 13-29.

Bhogapurapu, N., Dey, S., Bhattacharya, A., Mandal, D., Lopez-Sanchez, J.M., McNairn, H., López-Martínez, C. and Rao, Y.S., 2021. Dual-polarimetric descriptors from Sentinel-1 GRD SAR data for crop growth assessment. *ISPRS Journal of Photogrammetry and Remote Sensing*, *178*, pp.20-35.

Blickensdörfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasmi, S. and 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*, p.112831.

Boryan, C., Yang, Z., Mueller, R. and Craig, M., 2011. Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto International*, *26*(5), pp.341-358.

Botev, Z.I., Kroese, D.P., Rubinstein, R.Y. and L'Ecuyer, P., 2013. The cross-entropy method for optimization. In *Handbook of Statistics*, *31*, pp.35-59.

Breiman, L., 1996. Bagging predictors. *Machine learning*, *24*, pp.123-140.

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B. and Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sensing of Environment*, *210*, pp.35-47.

Chambers, L.G., 1989. 73.52 Spearman's rank correlation coefficient. *The Mathematical Gazette*, *73*(466), pp.331-332.

Chlingaryan, A., Sukkarieh, S. and Whelan, B., 2018. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, *151*, pp.61-69.

Cloude, S.R. and Pottier, E., 1997. An entropy based classification scheme for land applications of polarimetric SAR. *IEEE Transactions on Geoscience and Remote Sensing*, *35*(1), pp.68-78.

Dasari, K. and Lokam, A., 2018. Exploring the capability of compact polarimetry (hybrid pol) c band RISAT-1 data for land cover classification. *IEEE Access*, *6*, pp.57981-57993.

De, S., Kumar, V. and Rao, Y.S., 2014, June. Crop classification using RISAT-1 hybrid polarimetric SAR data. In *EUSAR 2014; 10th European Conference on Synthetic Aperture Radar*, pp.1-4.

Dingle Robertson, L., McNairn, H., Jiao, X., McNairn, C. and Ihuoma, S.O., 2022. Monitoring crops using compact polarimetry and the RADARSAT constellation mission. *Canadian Journal of Remote Sensing*, *48*(6), pp.793-813.

Dong, H., Xu, X., Wang, L. and Pu, F., 2018. Gaofen-3 PolSAR image classification via XGBoost and polarimetric spatial information. *Sensors*, *18*(2), p.611.

Dong, J., Xiao, X., Kou, W., Qin, Y., Zhang, G., Li, L., Jin, C., Zhou, Y., Wang, J., Biradar, C. and Liu, J., 2015. Tracking the dynamics of paddy rice planting area in 1986–2010 through time series Landsat images and phenology-based algorithms. *Remote Sensing of Environment*, *160*, pp.99-113.

Dong, J., Xiao, X., Zhang, G., Menarguez, M.A., Choi, C.Y., Qin, Y., Luo, P., Zhang, Y. and Moore, B., 2016. Northward expansion of paddy rice in northeastern Asia during 2000–2014. *Geophysical Research Letters*, *43*(8), pp.3754-3761.

Dou, P., Shen, H., Li, Z. and Guan, X., 2021. Time series remote sensing image classification framework using combination of deep learning and multiple classifiers system. *International Journal of Applied Earth Observation and Geoinformation*, *103*, p.102477.

Fisette, T., Rollin, P., Aly, Z., Campbell, L., Daneshfar, B., Filyer, P., Smith, A., Davidson, A., Shang, J. and Jarvis, I., 2013, August. AAFC annual crop inventory. In *2013 Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, pp.270-274.

Gao, H., Wang, C., Wang, G., Zhu, J., Tang, Y., Shen, P. and Zhu, Z., 2018. A crop classification method integrating GF-3 PolSAR and Sentinel-2A optical data in the Dongting Lake Basin. *Sensors*, *18*(9), p.3139.

Gómez, C., White, J.C. and Wulder, M.A., 2016. Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *116*, pp.55-72.

Griffiths, P., Nendel, C. and Hostert, P., 2019. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sensing of Environment*, *220*, pp.135-151.

Guo, J., Wei, P.L., Liu, J., Jin, B., Su, B.F. and Zhou, Z.S., 2018. Crop Classification Based on Differential Characteristics of $ H/\alpha $ Scattering Parameters for Multitemporal Quad-and Dual-Polarization SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(10), pp.6111-6123.

Hamad, R.A., Yang, L., Woo, W.L. and Wei, B., 2020. Joint learning of temporal models to handle imbalanced data for human activity recognition. *Applied Sciences*, *10*(15), p.5293.

He, C., He, B., Tu, M., Wang, Y., Qu, T., Wang, D. and Liao, M., 2020. Fully convolutional networks and a manifold graph embedding-based algorithm for polsar image classification. *Remote Sensing*, *12*(9), p.1467.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.770-778.

Heihe Social and Economic Statistics Yearbook. 2018. *Heihe Social and Economic Statistics Yearbook*. Beijing: China Statistical Publishing House.

Heine, I., Jagdhuber, T. and Itzerott, S., 2016. Classification and monitoring of reed belts using dual-polarimetric TerraSAR-X time series. *Remote Sensing*, *8*(7), p.552.

Hosseini, M., Becker-Reshef, I., Sahajpal, R., Lafluf, P., Leale, G., Puricelli, E., Skakun, S. and McNairn, H., 2022. Soybean Yield Forecast Using Dual-Polarimetric C-Band Synthetic Aperture Radar. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *3*, pp.405-410.

Inglada, J., Vincent, A., Arias, M. and Marais-Sicre, C., 2016. Improved early crop type identification by joint use of high temporal resolution SAR and optical image time series. *Remote Sensing*, *8*(5), p.362.

Ioannidou, M., Koukos, A., Sitokonstantinou, V., Papoutsis, I. and Kontoes, C., 2022. Assessing the added value of Sentinel-1 PolSAR data for crop classification. *Remote Sensing*, *14*(22), p.5739.

Ji, S., Zhang, C., Xu, A., Shi, Y. and Duan, Y., 2018. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sensing*, *10*(1), p.75.

Jia, Y., Jin, S., Savi, P., Gao, Y., Tang, J., Chen, Y. and Li, W., 2019. GNSS-R soil moisture retrieval based on a XGboost machine learning aided method: Performance and validation. *Remote Sensing*, *11*(14), p.1655.

Kamilaris, A. and Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, pp.70-90.

Kandasamy, S., Baret, F., Verger, A., Neveux, P. and Weiss, M., 2013. A comparison of methods for smoothing and gap filling time series of remote sensing observations–application to MODIS LAI products. *Biogeosciences*, *10*(6), pp.4055-4071.

Keller, C.A. and Evans, M.J., 2019. Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10. *Geoscientific Model Development*, *12*(3), pp.1209-1225.

Khan, H., Wang, X. and Liu, H., 2022. Handling missing data through deep convolutional neural network. *Information Sciences*, *595*, pp.278-293.

Khare, S., Bhandari, A., Singh, S. and Arora, A., 2012. ECG arrhythmia classification using spearman rank correlation and support vector machine. In *Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011), Springer, India*, *2*, pp.591-598.

King, L., Adusei, B., Stehman, S.V., Potapov, P.V., Song, X.P., Krylov, A., Di Bella, C., Loveland, T.R., Johnson, D.M. and Hansen, M.C., 2017. A multi-resolution approach to national-scale cultivated area estimation of soybean. *Remote Sensing of Environment*, *195*, pp.13-29.

Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M. and Inman, D.J., 2021. 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, *151*, p.107398.

Kumar, D., Rao, S. and Sharma, J.R., 2013, September. Radar Vegetation Index as an alternative to NDVI for monitoring of soyabean and cotton. In *Proceedings of the XXXIII INCA International Congress (Indian Cartographer), Jodhpur, India*, pp.19-21.

Kursa, M.B. and Rudnicki, W.R., 2010. Feature selection with the Boruta package. *Journal of statistical software*, *36*, pp.1-13.

Kwak, G.H., Park, C.W., Lee, K.D., Na, S.I., Ahn, H.Y. and Park, N.W., 2021. Potential of hybrid CNN-RF model for early crop mapping with limited input data. *Remote Sensing*, *13*(9), p.1629.

Li, H., Zhang, C., Zhang, S. and Atkinson, P.M., 2019. A hybrid OSVM-OCNN method for crop classification from fine spatial resolution remotely sensed imagery. *Remote Sensing*, *11*(20), p.2370.

Li, H., Zhang, C., Zhang, S. and Atkinson, P.M., 2020. Crop classification from full-year fully-polarimetric L-band UAVSAR time-series using the Random Forest algorithm. *International Journal of Applied Earth Observation and Geoinformation*, *87*, p.102032.

Liao, C., Wang, J., Xie, Q., Baz, A.A., Huang, X., Shang, J. and He, Y., 2020. Synergistic use of multi-temporal RADARSAT-2 and VENμS data for crop classification based on 1D convolutional neural network. *Remote Sensing*, *12*(5), p.832.

Mahdianpari, M., Mohammadimanesh, F., McNairn, H., Davidson, A., Rezaee, M., Salehi, B. and Homayouni, S., 2019. Mid-season crop classification using dual-, compact-, and full-polarization in preparation for the Radarsat Constellation Mission (RCM). *Remote Sensing*, *11*(13), p.1582.

Mahdianpari, M., Salehi, B. and Mohammadimanesh, F., 2017. The effect of PolSAR image de-speckling on wetland classification: introducing a new adaptive method. *Canadian Journal of Remote Sensing*, *43*(5), pp.485-503.

Mandal, D., Kumar, V., Ratha, D., Dey, S., Bhattacharya, A., Lopez-Sanchez, J.M., McNairn, H. and Rao, Y.S., 2020. Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data. *Remote Sensing of Environment*, *247*, p.111954.

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

McNairn, H., Shang, J., Jiao, X. and Champagne, C., 2009. The contribution of ALOS PALSAR multipolarization and polarimetric data to crop classification. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(12), pp.3981-3992.

Meier, U., Bleiholder, H., Buhr, L., Feller, C., Hack, H., Heß, M., Lancashire, P.D., Schnock, U., Stauß, R., Van Den Boom, T. and Weber, E., 2009. The BBCH system to coding the phenological growth stages of plants–history and publications. *Journal für Kulturpflanzen*, *61*(2), pp.41-52.

Mercier, A., Betbeder, J., Baudry, J., Le Roux, V., Spicher, F., Lacoux, J., Roger, D. and Hubert-Moy, L., 2020. Evaluation of Sentinel-1 & 2 time series for predicting wheat and rapeseed phenological stages. *ISPRS Journal of Photogrammetry and Remote Sensing*, *163*, pp.231-256.

Moumni, A. and Lahrouni, A., 2021. Machine learning-based classification for crop-type mapping using the fusion of high-resolution satellite imagery in a semiarid area. *Scientifica*, *2021*.

Mullissa, A.G., Persello, C. and Tolpekin, V., 2018, July. Fully convolutional networks for multi-temporal SAR image classification. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp.6635-6638.

Nair, V. and Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp.807-814.

Nasirzadehdizaji, R., Balik Sanli, F., Abdikan, S., Cakir, Z., Sekertekin, A. and Ustuner, M., 2019. Sensitivity analysis of multi-temporal Sentinel-1 SAR parameters to crop height and canopy coverage. *Applied Sciences*, *9*(4), p.655.

Navarro, A., Rolim, J., Miguel, I., Catalão, J., Silva, J., Painho, M. and Vekerdy, Z., 2016. Crop monitoring based on SPOT-5 Take-5 and sentinel-1A data for the estimation of crop water requirements. *Remote Sensing*, *8*(6), p.525.

Ndikumana, E., Ho Tong Minh, D., Baghdadi, N., Courault, D. and Hossard, L., 2018. Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sensing*, *10*(8), p.1217.

Nord, M.E., Ainsworth, T.L., Lee, J.S. and Stacy, N.J., 2008. Comparison of compact polarimetric synthetic aperture radar modes. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(1), pp.174-188.

Pal, M., 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), pp.217-222.

Pelletier, C., Webb, G.I. and Petitjean, F., 2019. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, *11*(5), p.523.

Qu, Y., Zhao, W., Yuan, Z. and Chen, J., 2020. Crop mapping from sentinel-1 polarimetric time-series with a deep neural network. *Remote Sensing*, *12*(15), p.2493.

Raney, R., 2006, July. Hybrid-polarity SAR architecture. In *2006 IEEE International Symposium on Geoscience and Remote Sensing*, pp.3846-3848.

Raney, R.K., 2019. Hybrid dual-polarization synthetic aperture radar. *Remote Sensing*, *11*(13), p.1521.

Raney, R.K., Cahill, J.T., Patterson, G.W. and Bussey, D.B.J., 2012. The m-chi decomposition of hybrid dual-polarimetric radar data with application to lunar craters. *Journal of Geophysical Research: Planets*, *117*(E12).

Rußwurm, M. and Körner, M., 2020. Self-attention for raw optical satellite time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *169*, pp.421-435.

Salma, S., Keerthana, N. and Dodamani, B.M., 2022. Target decomposition using dual-polarization sentinel-1 SAR data: Study on crop growth analysis. *Remote Sensing Applications: Society and Environment*, *28*, p.100854.

Siachalou, S., Mallinis, G. and Tsakiri-Strati, M., 2015. A hidden Markov models approach for crop classification: Linking crop phenology to time series of multi-sensor remote sensing data. *Remote Sensing*, *7*(4), pp.3633-3650.

Sonobe, R., 2019. Parcel-based crop classification using multi-temporal TerraSAR-X dual polarimetric data. *Remote Sensing*, *11*(10), p.1148.

Souyris, J.C., Imbo, P., Fjortoft, R., Mingot, S. and Lee, J.S., 2005. Compact polarimetry based on symmetry properties of geophysical media: The/spl pi//4 mode. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(3), pp.634-646.

Steele-Dunne, S.C., McNairn, H., Monsivais-Huertero, A., Judge, J., Liu, P.W. and Papathanassiou, K., 2017. Radar remote sensing of agricultural canopies: A review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(5), pp.2249-2273.

Sun, C., Bian, Y., Zhou, T. and Pan, J., 2019. Using of multi-source and multi-temporal remote sensing data improves crop-type mapping in the subtropical agriculture region. *Sensors*, *19*(10), p.2401.

Szantoi, Z., Escobedo, F., Abd-Elrahman, A., Smith, S. and Pearlstine, L., 2013. Analyzing fine-scale wetland composition using high resolution imagery and texture features. *International Journal of Applied Earth Observation and Geoinformation*, *23*, pp.204-212.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-9.

Teimouri, M., Mokhtarzade, M., Baghdadi, N. and Heipke, C., 2022. Fusion of time-series optical and SAR images using 3D convolutional neural networks for crop classification. *Geocarto International*, *37*(27), pp.15143-15160.

Teimouri, N., Dyrmann, M. and Jørgensen, R.N., 2019. A novel spatio-temporal FCN-LSTM network for recognizing various crop types using multi-temporal radar images. *Remote Sensing*, *11*(8), p.990.

Thenkabail, P.S., Knox, J.W., Ozdogan, M., Gumma, M.K., Congalton, R.G., Wu, Z.T., Milesi, C., Finkral, A., Marshall, M., Mariotto, I. and You, S., 2012. Assessing future risks to agricultural productivity, water resources and food security: How can remote sensing help?. *PE&RS, Photogrammetric Engineering & Remote Sensing*, *78*(8), pp.773-782.

Tomaszewski, M., Gasz, R. and Smykała, K., 2021. Monitoring Vegetation Changes Using Satellite Imaging–NDVI and RVI4S1 Indicators. In *Control, Computer Engineering and Neuroscience: Proceedings of IC Brain Computer Interface 2021*, pp. 268-278.

Ullmann, T., Schmitt, A., Roth, A., Duffe, J., Dech, S., Hubberten, H.W. and Baumhauer, R., 2014. Land cover characterization and classification of arctic tundra environments by means of polarized synthetic aperture X-and C-Band Radar (PolSAR) and Landsat 8 multispectral imagery—Richards Island, Canada. *Remote Sensing*, *6*(9), pp.8565-8593.

USDA-NASS, C. D. L., 2022. USDA National Agricultural Statistics Service Cropland Data Layer. Published crop-specific data layer. Available at: (https://nassgeodata.gmu.edu/CropScape/).

Valcarce-Diñeiro, R., Lopez-Sanchez, J.M., Sánchez, N., Arias-Pérez, B. and Martínez-Fernández, J., 2018. Influence of incidence angle in the correlation of C-band polarimetric parameters with biophysical variables of rain-fed crops. *Canadian Journal of Remote Sensing*, *44*(6), pp.643-659.

Van der Maaten, L. and Hinton, G., 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(11).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

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

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J.P., Camps-Valls, G. and Moreno, J., 2012. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and-3. *Remote Sensing of Environment*, *118*, pp.127-139.

Wang, L., Dong, Q., Yang, L., Gao, J. and Liu, J., 2019. Crop classification based on a novel feature filtering and enhancement method. *Remote Sensing*, *11*(4), p.455.

Wei, S., Zhang, H., Wang, C., Wang, Y. and Xu, L., 2019. Multi-temporal SAR data large-scale crop mapping based on U-Net model. *Remote Sensing*, *11*(1), p.68.

Wu, X., Xiao, X., Yang, Z., Wang, J., Steiner, J. and Bajgain, R., 2021. Spatial-temporal dynamics of maize and soybean planted area, harvested area, gross primary production, and grain production in the Contiguous United States during 2008-2018. *Agricultural and Forest Meteorology*, *297*, p.108240.

Xie, Q., Wang, J., Liao, C., Shang, J., Lopez-Sanchez, J.M., Fu, H. and Liu, X., 2019. On the use of Neumann decomposition for crop classification using multi-temporal RADARSAT-2 polarimetric SAR data. *Remote Sensing*, *11*(7), p.776.

Xu, J., Yang, J., Xiong, X., Li, H., Huang, J., Ting, K.C., Ying, Y. and Lin, T., 2021. Towards interpreting multi-temporal deep learning models in crop mapping. *Remote Sensing of Environment*, *264*, p.112599.

Xu, J., Zhu, Y., Zhong, R., Lin, Z., Xu, J., Jiang, H., Huang, J., Li, H. and Lin, T., 2020. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. *Remote Sensing of Environment*, *247*, p. 111946.

Xu, L., Zhang, H., Wang, C., Zhang, B. and Liu, M., 2018. Crop classification based on temporal information using sentinel-1 SAR time-series data. *Remote Sensing*, *11*(1), p.53.

Yang, L., Wang, L., Huang, J., Mansaray, L.R. and Mijiti, R., 2019. Monitoring policy-driven crop area adjustments in northeast China using Landsat-8 imagery. *International Journal of Applied Earth Observation and Geoinformation*, *82*, p.101892.

Yang, Q., Yuan, Q., Gao, M. and Li, T., 2023. A new perspective to satellite-based retrieval of ground-level air pollution: Simultaneous estimation of multiple pollutants based on physics-informed multi-task learning. *Science of The Total Environment*, *857*, p.159542.

Yang, S., Gu, L., Li, X., Jiang, T. and Ren, R., 2020. Crop classification method based on optimal feature selection and hybrid CNN-RF networks for multi-temporal remote sensing imagery. *Remote Sensing*, *12*(19), p.3119.

You, N. and Dong, J., 2020. Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, *161*, pp.109-123.

You, N., Dong, J., Huang, J., Du, G., Zhang, G., He, Y., Yang, T., Di, Y. and Xiao, X., 2021. The 10-m crop type maps in Northeast China during 2017–2019. *Scientific Data*, *8*(1), p.41.

Yuan, Y., Lin, L., Huo, L.Z., Kong, Y.L., Zhou, Z.G., Wu, B. and Jia, Y., 2020. Using an attention-based LSTM encoder–decoder network for near real-time disturbance detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, pp.1819-1832.

Zhang, C., Di, L., Lin, L., Li, H., Guo, L., Yang, Z., Eugene, G.Y., Di, Y. and Yang, A., 2022. Towards automation of in-season crop type mapping using spatiotemporal crop information and remote sensing data. *Agricultural Systems*, *201*, p.103462.

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J. and Atkinson, P.M., 2019. Joint Deep Learning for land cover and land use classification. *Remote Sensing of Environment*, *221*, pp.173-187.

Zhang, G., Xiao, X., Dong, J., Kou, W., Jin, C., Qin, Y., Zhou, Y., Wang, J., Menarguez, M.A. and Biradar, C., 2015. Mapping paddy rice planting areas through time series analysis of MODIS land surface temperature and vegetation index data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *106*, pp.157-171.

Zhang, Q., Liu, Y., Gong, C., Chen, Y. and Yu, H., 2020. Applications of deep learning for dense scenes analysis in agriculture: A review. *Sensors*, *20*(5), p.1520.

Zhang, S., 2012. Nearest neighbor selection for iteratively kNN imputation. *Journal of Systems and Software*, *85*(11), pp.2541-2552.

Zhang, X., Dierking, W., Zhang, J. and Meng, J., 2014. A polarimetric decomposition method for ice in the Bohai Sea using C-band PolSAR data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(1), pp.47-66.

Zhao, H., Chen, Z., Jiang, H., Jing, W., Sun, L. and Feng, M., 2019. Evaluation of three deep learning models for early crop classification using sentinel-1A imagery time series—A case study in Zhanjiang, China. *Remote Sensing*, *11*(22), p.2673.

Zhong, L., Gong, P. and Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. *Remote Sensing of Environment*, *140*, pp.1-13.

Zhong, L., Hu, L. and Zhou, H., 2019. Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, *221*, pp.430-443.

Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H. and Xu, B., 2016, August. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 2, pp.207-212.

Zhou, T., Pan, J., Zhang, P., Wei, S. and Han, T., 2017. Mapping winter wheat with multi-temporal SAR and optical images in an urban agricultural region. *Sensors*, *17*(6), p. 1210.

Zhou, Y.N., Luo, J., Feng, L. and Zhou, X., 2019. DCN-based spatial features for improving parcel-based crop classification using high-resolution optical images and multi-temporal SAR data. *Remote Sensing*, *11*(13), p.1619.

Zhou, Z.H., 2012. *Ensemble Methods: Foundations and Algorithms*. CRC press.

Zhu, X.X., Montazeri, S., Ali, M., Hua, Y., Wang, Y., Mou, L., Shi, Y., Xu, F. and Bamler, R., 2021. Deep learning meets SAR: Concepts, models, pitfalls, and perspectives. *IEEE Geoscience and Remote Sensing Magazine*, *9*(4), pp.143-172.

Chapter 4 Enhanced Crop Classification through Integrated Optical and SAR Data: A Deep Learning Approach for Multi-Source Image Fusion

This chapter is based on the published paper by Liu, N., Zhao, Q., Williams, R. and Barrett, B., 2023. Enhanced crop classification through integrated optical and SAR data: a deep learning approach for multi-source image fusion. *International Journal of Remote Sensing*, pp.1-29, doi: 10.1080/01431161.2023.2232552.

Abstract

Agricultural crop mapping has advanced over the last decades due to improved approaches and the increased availability of image datasets at various spatial and temporal resolutions. Considering the spatial and temporal dynamics of different crops during a growing season, multi-temporal classification frameworks are well-suited for mapping crops at large scales. Addressing the challenges posed by imbalanced class distribution, our approach combines the strengths of different deep learning models in an ensemble learning framework, enabling more accurate and robust classification by capitalizing on their complementary capabilities. This research aims to enhance the crop classification of maize, soybean, and wheat in Bei'an County, Northeast China, by developing a novel deep learning architecture that combines a three-dimensional convolutional neural network (3D-CNN) with a variant of convolutional recurrent neural networks (ConvRNN). The proposed method integrates multi-temporal Sentinel-1 polarimetric features with Sentinel-2 surface reflectance data for multi-source fusion and achieves an overall accuracy of 91.7%, a Kappa coefficient of 85.7%, and F1 scores of 93.7%, 92.2%, and 90.9% for maize, soybean, and wheat, respectively. Our proposed model is also compared with alternative data augmentation techniques, maintaining the highest mean F1 score (87.7%). The best performer was weakly supervised with ten per cent of ground truth data collected in Bei'an in 2017 and used to produce an annual crop map for measuring the model's generalisability. The model learning reliability of the proposed method is interpreted through the visualisation of model soft outputs and saliency maps.

Keywords: Agricultural crop mapping, multi-temporal classification, deep learning, 3D-CNN, ConvRNN, Multi-source image fusion

4.1 Introduction

Crop mapping is essential for the assessment of the underlying factors for farming system changes and the management of crops. Northeast China has become one of the main breadbaskets of the country, serving an increasingly important role in agricultural production and international trade of certain crops such as soybeans (Dong et al., 2016; Yang et al., 2019). Targeting the economic sustainability of agricultural development, however, the retrieval of quantitative information from the changes in the local croplands has been limited due to the annual crop rotation practice featured in this region (You et al., 2021). As such, accurate annual crop maps are still in high demand by local authorities in China to build near real-time crop monitoring mechanisms for early yield assessment of major crops at the county-level scale. Many studies have made considerable progress in the development of crop mapping systems by using satellite imagery with moderate spatial resolutions due to their coverage and regular repeat acquisitions (Boryan et al., 2011; Inglada et al., 2015; Defourny et al., 2019). Considering the spectral characteristics observed in commonly used optical satellite sensors such as Landsat, MODIS and Sentinel-2, many studies have investigated and quantified the dynamics (i.e., seasonal changes) of vegetation indices (VIs) and optical bands, using them as distinctive input features to accurately identify crop types throughout the growing seasons. (Fan et al., 2014; Zheng et al., 2015; Zhong et al., 2016a; Zhong et al., 2016b; Song et al., 2017; You and Dong, 2020). In light of existing research in automated crop identification, our study seeks to develop a novel approach for enhancing crop mapping performance leveraging the potential of satellite remote sensing data, which can contribute toward addressing the pressing need for sustainable agricultural development in Northeast China.

Cloud cover and/or adverse weather conditions can limit the quality of optical acquisitions and impact upon crop monitoring capabilities, resulting in data loss within time series of satellite acquisitions during the growing season (Sonobe et al., 2014; Kussul et al., 2018; Griffiths et al., 2019). Synthetic aperture radar (SAR) sensors are active remote sensors that can operate independently of weather conditions or solar illumination. SAR images provide unique radar-related information primarily responding to the biophysical properties of vegetation (e.g., Gao et al., 2018; Sun et al., 2019; Qu et al., 2020). Many studies have demonstrated the feasibility of using radar polarimetric features to detect crop types, generated by specific polarimetric decomposition algorithms that include Pauli, Cloude-Pottier, Freeman-Durden, H/A/α, Huynen, Yamaguchi Neumann and Krogager (He et al., 2020; Liao et al., 2020; Xie et al., 2019; Gao et

al., 2018). Both optical and radar data have respectively demonstrated their capability for crop mapping, and the fusion of image data from both data sources is increasingly explored to improve the crop mapping performance (e.g. Gao et al., 2018; Liao et al., 2020; Moumni and Lahrouni, 2021; Sun et al., 2019; Van Tricht et al., 2018; Li et al., 2022). The combination of optical and radar data provides complementary information that can reduce temporal gaps in data capture, which can contribute significantly to identifying crops in cloud-prone regions (Sun et al., 2019; Liao et al., 2020). Similarly, combining multi-sensor data yields richer information on certain crops to overcome the heterogeneity of some areas caused by mixed crops (Moumni and Lahrouni, 2021). Most previous studies have either stacked optical and radar data at the pixel level for crop classification (e.g. Gao et al., 2018; Liao et al., 2020; Moumni and Lahrouni, 2021; Van Tricht et al., 2018), or independently trained the image data from dual sources, e.g., Sentinel-1 and Sentinel-2, using separate models in parallel. Subsequently, the resultant outputs from each model are integrated into one learned feature sequence (Teimouri et al., 2022).

Previous studies employed machine learning models, such as Decision Tree (DT), Support Vector Machine (SVM) and Random Forest (RF) to identify crops based on multi-temporal observations (Zhong et al., 2014; Pelletier et al., 2016; Bargiel 2017; Teluguntla et al., 2018 Gao et al., 2018), however conventional machine learning models were not originally designed to process temporal data. Additionally, the enhanced representation of crop growth patterns requires phenological metrics defined with expertise in multi-temporal remote sensing data (You and Dong, 2020), and those designed metrics are not always available until the end of the crop growth cycle (Xu et al., 2021). Although machine learning approaches improve classification performance with increasing dimensions of input variables and reduce the requirements for designating threshold-based classification rules, the temporal relationship in multi-temporal satellite data cannot be fully and automatically utilised. More recently, studies demonstrated that a series of deep learning networks could successfully explore the sequential relationships within time-series remote sensing data for crop classification (Crisóstomo de Castro Filho et al., 2020; Dou et al., 2021; Liao et al., 2020; Rußwurm and Körner 2020; Sun et al., 2020; Xu et al., 2020; Zhao et al., 2021; Zhong et al., 2019). These deep neuron-based architectures include one-dimensional Convolutional Neural Networks (1D-CNNs), Long Short-Term Memory (LSTM) and variants or combinations of both architectures. Given that these architectures by the 1D-CNN or LSTM models are naturally fitted with extracting sequential dependencies within multi-temporal remote sensing data, these models generally

outperform the nontemporal models such as RF in terms of classification performance for maize and soybean (Xu et al., 2020) and other crops (Liao et al., 2020; Rußwurm and Körner 2020; Zhong et al., 2019). However, temporal models are not used for the extraction of spatial features from satellite imageries.

The spatial relationship, known as the spatial arrangement of the adjacent pixels represented by the data matrix in remote sensing images, is also a main consideration for crop classification with remote sensing data. Two-dimensional Convolutional Neural Networks (2D-CNNs) are used to extract multi-level spatial features from satellite data for crop classification (Kussul et al., 2017; Wei et al., 2019; He et al., 2020). A patch-based CNN architecture is designed for regional-level classification on medium-resolution satellite imagery by collecting a series of image patches as inputs instead of pixel-based samples used for machine learning models, 1D-CNNs or LSTM models (Sharma et al., 2017). 2D-CNNs only focus on the spatial dimension due to the multidimensional input (the image size and the channel-wise image bands), whereas the temporal dependencies are not considered. Therefore, three-dimensional Convolutional Neural Networks (3D-CNNs) are proposed for the extraction of spatiotemporal features from image data. Fewer studies have applied 3D-CNN-based architectures for crop classification (e.g. Adrian et al., 2021; Ji et al., 2018; Teimouri et al., 2022). Roy et al. (2019) showed that a hybrid 3D-2D CNN had an improved performance over using standalone 3D-CNN and 2D-CNN, respectively. Another approach to obtaining spatiotemporal features is Convolutional Recurrent Neural Networks (ConvRNNs), and the variants represented by different recurrent units have been used to identify a large number of crop classes in a hierarchical framework (Turkoglu et al., 2021a). To the best of our knowledge, there is less research regarding the synergistic use of 3D-CNN, 2D-CNN and ConvRNN architectures for crop classification.

Despite the findings in previous studies, annual crop mapping in Northeast China remains challenging due to the high intra-class variance and inter-class similarity of spectral qualities and phenology of crops in the region, which are influenced by varying climate conditions, geomorphic characteristics, and cropping systems (Wang et al., 2019). Additionally, regular and cloud-free time series acquisitions are often limited to agriculture monitoring at a large scale (Defourny et al., 2019). As a result, this study utilizes a small number of available optical acquisitions for large-scale crop mapping as supplementary sources for time series SAR data to develop models that enhance crop mapping accuracy. The study aims to develop a novel framework that combines 3D-CNN, 2D-CNN, and ConvRNN architectures for county-level

crop mapping based on the fusion of multi-temporal optical and SAR images for Bei'an county in Northeast China in 2017 at a 10 m spatial resolution. This spatiotemporal model contributes to improved performance in identifying crops during the growing season and addressing imbalanced class distribution, which could lead to model bias towards majority classes. The resulting crop maps can be used for dynamic monitoring of interannual crop growth in the same area and provide annual crop inventory information for local authorities to evaluate land-use policies. In this study, the proposed model is assessed for crop mapping and juxtaposed with models presented in previous studies (Pelletier et al., 2019; Ji et al., 2018; Turkoglu et al., 2021b; Roy et al., 2019). Additionally, the models are examined in relation to data augmentation techniques and evaluated across three randomly selected geographical locations. Subsequently, the optimally chosen model is employed to generate an annual crop map for Bei'an in 2017 through model inference.

4.2 Study Area

Bei'an is a county located in the northeast part of Heilongjiang province in China (47°35'N ~ 48°33'N, 126°16'E ~ 127°53'E) (Figure 4-1). According to Bei'an Municipal People's Government (http://www.hljba.gov.cn/), the total area of Bei'an county is approximately 7149 km^2. Bei'an is subject to a cold and temperate continental monsoon climate. The average annual temperature is around 1.2 °C with annual effective accumulated temperature ranging from 18.30 °C to 23.50 °C. Bei'an receives an average annual precipitation of 529 millimetres, with the majority of rainfall occurring during the summer months from June to August. The average total amount of annual surface water resources is approximately 1.156 billion cubic meters. Bei'an is geographically located in the transitional zone between Songnen Plain and the Khingan Mountains, which is regarded as one of the world's three Chernozem (black soil) belts. Given the favourable soil fertility, meteorological conditions and regional temperature, this region serves as an ideal ecological habitat conducive to crop growth and agricultural yield. According to Heihe Social and Economic Statistics Yearbook (2018), the total crop sown area of Bei'an approximates 2190 km^2. Summer maize and soybean are the primary crop types, accounting for 29.5% and 61.8% of the total sowing area, respectively. In contrast, wheat, as one of the minority crop types in Bei'an, covers 2.9% of the total sown area. According to the local crop sowing scheme, the growing season of maize often spans from late April to late September, and soybeans are normally sown from early May to mid-September. These periods might vary annually due to crop rotation cycles in the study area over the years.

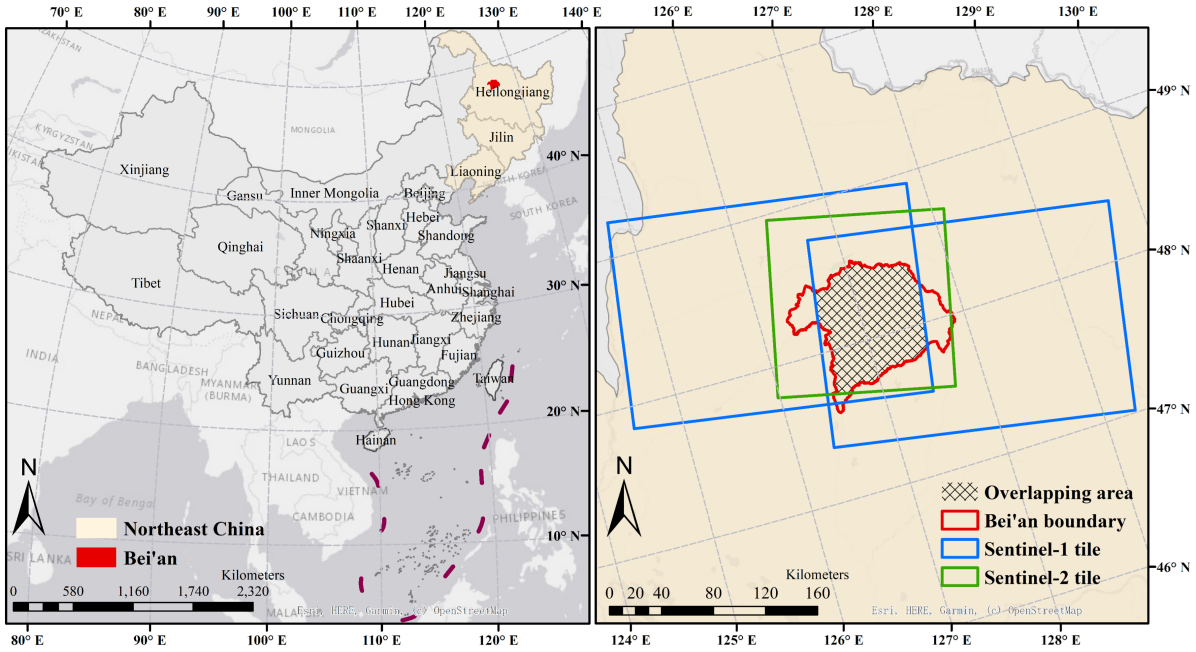

Figure 4-1. The study area in Bei'an. The multi-temporal Sentinel-1 and Sentinel-2 data are overlapped to capture the area that is covered by complete time-series acquisitions.

4.3 Datasets

4.3.1 Sentinel-1/2 datasets and pre-processing

In this study, both time-series Sentinel-1B Single Look Complex products (Interferometric Wide swath SLC) and Sentinel-2A/B (Level-1C) image datasets were acquired from the Sentinel Scientific Data Hub (https://scihub.copernicus.eu/dhus/#/home). Considering the local cropping practice in which the majority of crops were planted and harvested from early May to late September 2017, the image acquisitions were collected from 6th May 2017 to 27th September 2017, corresponding to the vegetative growing cycle of the recorded staple crops in Bei'an. As such, twenty-three Sentinel-1 acquisitions and three Sentinel-2 acquisitions were collected. The selection of Sentinel-2 data was based on the criteria that the average percentage of cloud coverage for the acquisition candidates is less than 8%.

The pre-processing of time series Sentinel-1 images was completed using the Sentinel Application Platform (SNAP) developed by the European Space Agency (ESA). The standard pre-processing steps follow Qu et al. (2020) which typically include radiometric calibration, multi-temporal speckle filtering (Refined Lee) and geocoding. Backscatter values were converted to decibel (dB) scale, and the cross-ratio of the backscatter was calculated by subtracting VV from VH, in accordance with logarithm rules. Sentinel-1 operates as an

inherently dual-polarised SAR platform, which can constrain the extent of polarimetric information that can be explored, compared to quad-polarimetric SAR systems providing fully polarimetric observations. However, quad-polarisation satellite acquisitions often suffer from reduced swath coverage, revisit time and accessibility. Hence, a compact polarimetric technique, m-chi decomposition (Raney, 2012), has demonstrated its utility for crop mapping using dual-pol data (Sonobe et al., 2019). The m-chi decomposition parameters were also obtained using SNAP. Each type of Sentinel-1 data was resampled to 10 m spatial resolution.

The pre-processing of Sentinel-2 images consists of the transformation from top-of-atmosphere (TOA) Sentinel-2 Level-1C reflectance images to bottom-of-atmosphere (BOA) Level-2A using Sen2Cor. Additionally, Band 4 (Red, 10 m), Band 8A (Vegetation Red Edge, 20 m), and Band 11 (SWIR, 20 m) were selected for their sensitivity to differentiate soybean and maize in Northeast China (You and Dou, 2021). All selected bands were resampled to 10 m and collocated with SAR data in a time-series sequence. Finally, a global min/max normalization approach was applied to all input features using the scikit-learn package to hasten the convergence of deep learning algorithms.

4.3.2 Ground truth and partitioning

Ground surveys of the study area were conducted during July, August and September 2017 by the Chinese Academy of Agricultural Sciences (CAAS). During the 2017 period, 21,257 fields were surveyed to calculate the area of crop parcels, record crop categories and retrieve annual statistics during the agricultural household survey. For cropland parcels with various crop types, they were manually digitised and labelled based on 5-meter resolution RapidEye satellite imagery (NIR Infra-red, Red Edge and Red composite), while Sentinel-2A images (SWIR, Narrow NIR and Red composite) were used for drawing relatively large cropland parcel areas with uniform crop types. In total, the classes of interest for major crops were assigned unique labels, including maize and soybean. In the in-situ dataset provided by CAAS, a small number of polygons were also identified for wheat and unknown crops. The proportions of ground sample pixels for each class in 2017 and the distribution of crop parcel size are displayed in Figure 4-2. A cropland mask layer, produced for Bei'an in 2017, is used to exclude non-cropland areas in this study during the model inference stage for the generation of an annual crop map. The cropland distribution and extent barely changed during 2017-2019 due to cropland protection by policies in Northeast China (Liu et al., 2014; Ning et al., 2018).

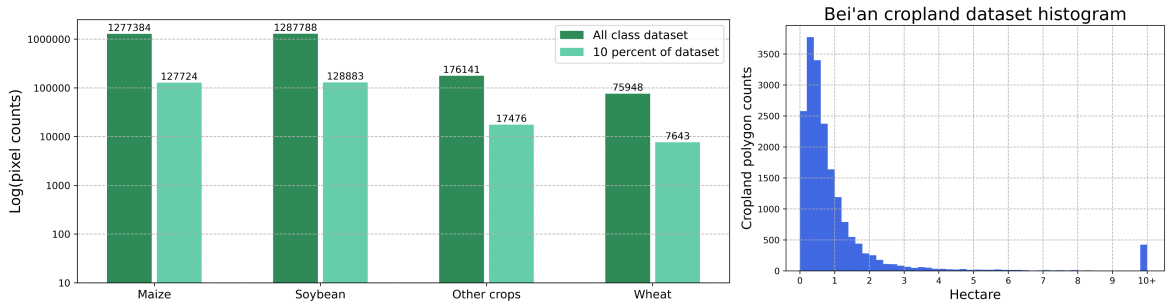

Figure 4-2. The sample class distribution with the number of pixels (y-axis) at the logarithmic scale for the Bei'an dataset collected in 2017 (left), and 10 percent of the dataset is split into subsets for training, validation, and testing (left). The distribution of crop parcel size overall (right). The parcels large than 9 hectares are accumulated in the last bin in the histogram. The parcel size on average is 1.39 hectares.

Since the cropping and managing system for each crop parcel would be different, the pixels within the same crop polygon are strongly correlated and need to be isolated when assigned to training, validation and testing data sets. i.e., pixels in each set should be mutually exclusive and not from the same crop parcels. Additionally, the class distributions in all sets should be identical (Rußwurm and Körner, 2017). In most croplands, pixels in the same parcel are very homogenous and highly correlated. Allocating pixels in a parcel to different sets will violate the principle of independence. The model generalisation on truly unseen data would be affected because it is likely that models have seen at least parts of the image patches used for validation (Audebert et al., 2019). Although the study area can be split into relatively large sub-regions, the crop types are not usually distributed evenly in the study area, which cannot ensure sub-regions with similar class distributions (Zhong et al., 2019). In this study, each crop polygon was regarded as an entity. The parcels are grouped using grids at 10-kilometre intervals so that crop parcels in the same grid are considered as a whole. The dataset is divided into training, validation, and testing partitions using image grids, which are selected at random in a ratio of 60%, 20%, and 20%, respectively. Additionally, 10% of the ground samples are randomly chosen from each selected set using stratified sampling, maintaining the same ratio as the initial dataset division.

4.4 Methods

4.4.1 Methodology framework

The entire workflow in this study is depicted in Figure 4-3, outlining the four stages designed to evaluate deep learning approaches for crop mapping using the fusion of multi-source, time-series satellite data. The initial stage focuses on the pre-processing of multi-temporal satellite acquisitions, specifically targeting the overlapping area in Bei'an (Figure 4-1). Section 4.3.1 describes the data pre-processing in detail. In the data preparation stage, three Sentinel-2 acquisitions with designated reflectance bands are stacked with Sentinel-1 intensity and polarimetric bands, respectively. This creates an accumulated time series dataset containing 26 satellite acquisitions arranged chronologically by the acquisition dates. This sequence covers the entire growth stages of the crops. The resultant dataset, characterized by band-wise stacking, includes 78 channels, covering both the collocated optical/backscatter and optical/polarimetric bands. During the initial data preparation stage, this stacked dataset is segmented into tiles using 10-kilometre grids, as explained in section 4.3.2. These image tiles are subsequently subdivided into smaller patches in batch processing, with each patch aligned with its corresponding ground truth label. Specifically, the centre pixel of each patch is directly associated with a specific crop label, and square patches representing ground samples (i.e., crop labels) are selected for input into the CNN models considered in this study. Section 4.4.4 will discuss the optimal size for these patches.

The experimental stage compares the performance of the proposed model with other state-of-the-art methods, given multiple model input scenarios. Particularly, an ablation experiment is conducted during model training and testing to determine the key input scenario for crop identification in Bei'an 2017. Following this, this study assesses the efficacy of implementing data augmentation techniques. The final stage involves generating a county-level crop map using the best performer and analysing the model learning outcomes. In the subsequent sections, the specifics of the experiment are introduced, presenting aspects such as classification algorithms employed, the environment in which the models are deployed, compact polarimetric parameters, and augmentation techniques, to provide a comprehensive understanding of the methodology in this study.

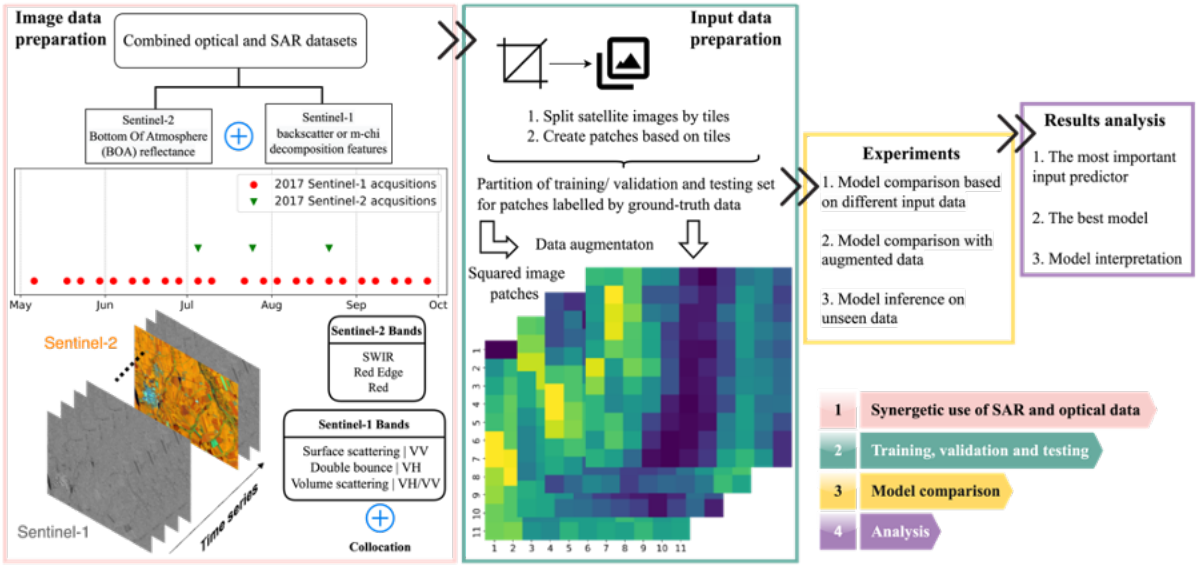

Figure 4-3. The overall workflow of the experiments.

4.4.2 Classification methods

4.4.2.1 3D-CNN

Convolutional Neural Networks (CNNs), motivated by the animal's visual cortex, is a deep learning technique to extract features in a way that considers spatial contexts between pixels instead of focusing on a single vector transformed by the typical multilayer neural networks (Sharma et al., 2017). CNNs, therefore, are also known as a dimensionality reduction method to handle multidimensional inputs in terms of its unique feature extraction pattern performed by convolutional kernels. Conventional two-dimensional CNNs (2D-CNN), however, are limited to the spatial features and may produce overwhelming parameters if the multidimensional inputs have large channels (spectral information) or time steps (temporal information) (Mäyrä et al., 2021). Conversely, the one-dimensional CNNs (1D-CNN) extract features from single-pixel temporal or spectral profiles of the input data without considering the spatial relationship between features. Although 2D-CNN can be combined with 1D-CNN to extract spatial-spectral or spatial-temporal information for improved results compared to dealing with information in only one dimension (Audebert et al., 2019), the large number of model parameters will be needed by 2D-CNN. An alternative method to extract features simultaneously on both dimensions is three-dimensional CNNs (3D-CNN). The convolutional kernels in 3D-CNN are cubes and produce a feature map with volume rather than a two-dimensional image derived by 2D-CNN or a single vector by 1D-CNN. The three-dimensional convolving process can be written as follows:

$$Y_{i,j}^{x,y,d} = Relu(\sum_n \sum_{t=0}^{T} \sum_{p=0}^{S} \sum_{q=0}^{S} w_{(i-1),n}^{t,p,q} X_{(i-1),n}^{(x+p),(y+q),(d+t)} + b_{i,j}) \quad (4\text{-}1)$$

Where $Y_{i,j}^{x,y,d}$ is the output values at 3D coordinate (x, y, d) on the j^{th} feature cube in the i^{th} layer, (x, y) is the spatial position and d indicates temporal index. $w_{(i-1),n}^{t,p,q}$ is the 3D kernel value at location (t, p, q) from the n^{th} feature cube in the previous layer. Similarly, (p, q) is the spatial position, and t denotes the temporal indicator of the kernel. $X_{i-1,n}^{x,y,d}$ is the input at position (x, y, d) from the n^{th} feature cube in the previous layer. $b_{i,j}$ is the bias vector on the j^{th} feature cube in the i^{th} layer. The size of the kernel is $T\times S\times S$ which is equivalent to length, height and width. Empirically, the size of height equals width in CNNs.

4.4.2.2 ConvSTAR

Convolutional recurrent neural network (ConvRNN) is a variant of sequence modelling that is built with convolutional operations in state transitions instead of matrix multiplications for handling spatio-temporal data. A typical convolutional Long Short-Term Memory (ConvLSTM) is designed to capture the spatio-temporal correlations for precipitation forecast (Shi et al., 2015), which outperforms the general LSTM structure in which spatial information is not considered. Given that stacking multiple ConvRNN layers contributes to feature extraction, a novel recurrent cell, namely STAckable Recurrent cell (STAR), is developed to reduce the exploding gradient effects and the number of trainable parameters (Turkoglu et al., 2021b). The convolutional version of STAR (ConvSTAR) is the modification of ConvLSTM in which the input and output gate are removed, which can be written as:

$$f_t = \sigma\,(W_{xf} * X_t + W_{hf} * H_{t-1} + B_f) \quad (4\text{-}2)$$

$$Z_t = tanh(W_{xz} * X_t + B_z) \quad (4\text{-}3)$$

$$H_t = \tanh\,(\,f_t \cdot\, Z_t + H_{t-1} \cdot (1 - f_t)) \quad (4\text{-}4)$$

Where σ is the sigmoid activation function, $*$ denotes the convolution operator and $\cdot$ indicates the Hadamard product (elementwise). The input X_t is firstly non-linearly projected through the activation function in Z_t. In addition, the previous state H_{t-1} and new inputs are linearly combined in the gating module $which$ is the determinant of the state-to-state flow to create a

new hidden state. W and B are matrices for weight and bias, respectively. The hidden state H_t is the output of a single ConvSTAR layer, which can be used for classification, or be used as the new inputs for the next layer or other decoders. Figure 4-4 illustrates a ConvSTAR cell that integrates Eq. (4-2), Eq. (4-3) and Eq. (4-4). The code of the ConvSTAR layer was adapted in Tensorflow format from the Pytorch repository (https://github.com/0zgur0/STAckable-Recurrent-network.git).

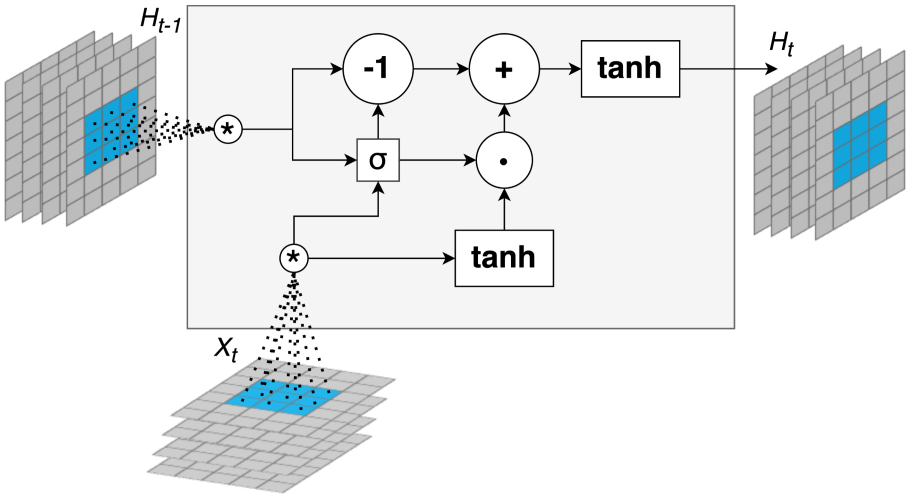

Figure 4-4. The structure of a ConvSTAR cell.

4.4.2.3 Synergic use of 3D-CNN and ConvSTAR

Recent studies have applied the combination of 3D-CNN and convolutional recurrent networks for univariant and multivariate time series forecasting. The 3D-CNN layers and attention ConvLSTM layers are utilised sequentially for multispectral soybean prediction (Nejad et al., 2022). The 3D-CNN layer can also be fed with features produced by ConvLSTM, serving as the model output layer for predicting urban expansion in image segmentation (Boulila et al., 2021). The enhanced performance on human action recognition was also proved by the combined use of these two deep learning approaches (Wang et al., 2021). In this study, a similar hybrid feature learning framework, 3D-ConvSTAR, is proposed to improve crop classification performance (Figure 4-5). The proposed network consists of three stages. The first step is made of a three-layer 3D-CNN with the optimal kernel size $3 \times 3 \times 3$, considering that three convolutional layers demonstrated effectiveness over two-layer and four-layer networks (Ji et al., 2018). Each layer has 32, 32, and 64 filters, respectively. The feature maps after each 3D convolutional operation are not shrunk by applying zero padding. The convolutional cubes are moved during one step. As previously introduced in Section 4.4.2.1, 3D-CNN is used to extract

spatio-temporal features simultaneously. Next, the output tensors from the 3D-CNN module are reshaped and fed into a three-layer bidirectional ConvSTAR unit. Bidirectional recurrent cells preserve temporal information from both the future and past, alleviating temporal bias toward data in later time steps (Rußwurm and Körner., 2018). The kernel size for ConvSTAR is 3×3 and the number of kernels for each layer is set to 64. Followed by ConvSTAR layers is a shallow 1-layer 2D-CNN with 64 3×3 kernels to take in the final hidden state from the previous layer for further extracting discriminative feature maps on the spatial dimension. It also can perform dimensionality reduction to some extent so that the number of model parameters can be optimised since it reduces the size of the feature maps and preserves the main information captured by the previous layers (Mäyrä et al., 2021; Roy et al., 2019). The last part of 3D-ConvSTAR is constructed with three fully connected (FC) layers. The last two dense layers have 256 and 128 units respectively and both are followed by a dropout layer with a factor of 0.4 to prevent networks from overfitting. The activation function Rectified Linear Unit (ReLU) (Nair and Hinton, 2010) is applied after CNN, convolutional recurrent and FC layers to augment the nonlinearity of outputs and control model systematic errors (bias). Pooling layers are not applied after all layers in the proposed model since it could cause loss of information at multiple dimensions (Li et al., 2019).

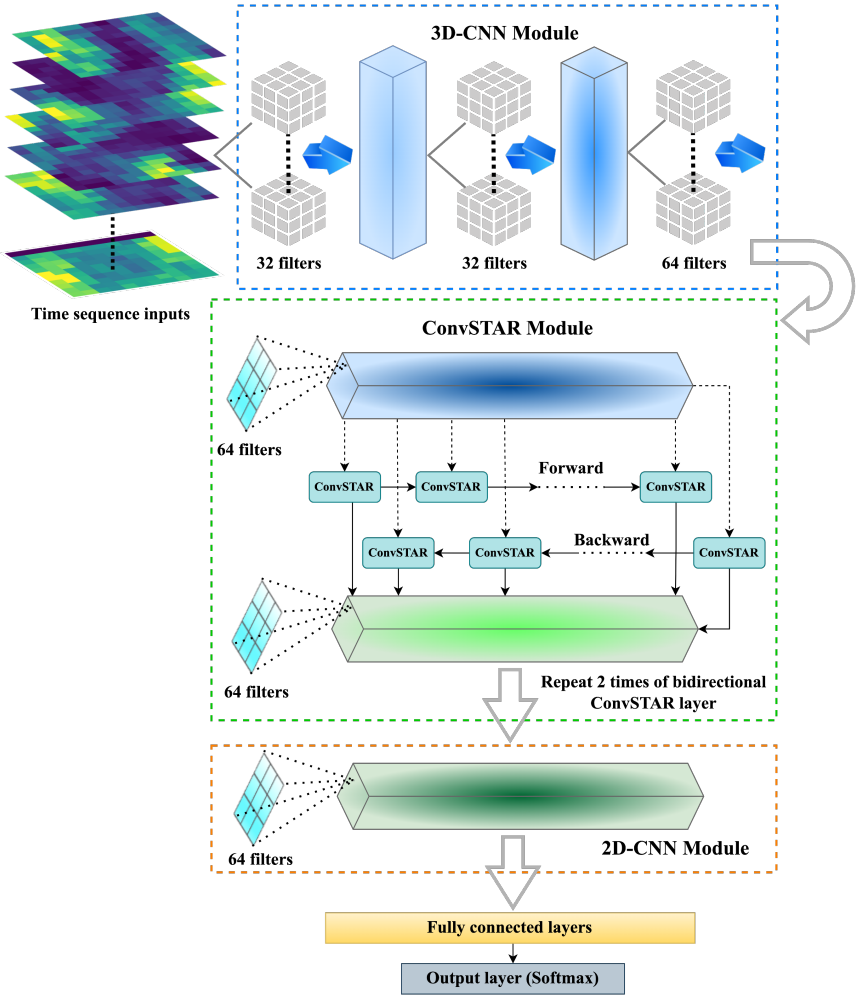

Figure 4-5. The architecture of 3D-ConvSTAR.

4.4.3 M-chi decomposition

A similar methodology was developed for single-transmitted dual-receive polarisation data that transmits at circular polarisation and receives at horizontal and vertical polarisation (Raney et al., 2012). This decomposition methodology, originally for compact polarimetric radar data, is based on the 2 × 2 covariance matrix, which is not applicable to general quad-pol data. This method is often characterised by the form composed of four-element Stokes parameters, which provides potential for hybrid polarimetric and dual-pol data. One application is the m-chi decomposition, in which the observed field is characterised by the degree of polarisation (m) as:

$$m = \frac{\sqrt{S_2^2+S_3^2+S_4^2}}{S_1}, \quad 0 \le m \le 1 \tag{4-5}$$

Where $S_{1,2,3,4}$ represents four Stokes parameters in the total power over an image field, and m refers to the degree of polarisation. $Chi\ (\chi)$, a Poincaré variable, denotes the field's ellipticity and circularity, which can be expressed through:

$$\sin 2\chi = -\frac{S_4}{mS_1}, \quad -45° \leq \chi \leq +45° \tag{4-6}$$

Based on these two variables calculated from the Stokes parameters, three target scattering parameters P_s (Single-bounce scattering), P_d (Double-bounce scattering) and P_v (Volume scattering) for m-chi decomposition can be expressed as follows:

$$P_s = \sqrt{\frac{mS_1(1-\sin 2\chi)}{2}} \tag{4-7}$$

$$P_d = \sqrt{\frac{mS_1(1+\sin 2\chi)}{2}} \tag{4-8}$$

$$P_v = \sqrt{S_1(1-m)} \tag{4-9}$$

Previous studies have shown that hybrid polarimetric data are close to and occasionally equivalent to the analysis of quad-pol data e.g., using the conventional Freeman-Durden decomposition (Nord et al., 2008; Ainsworth et al., 2009). Even though all required four-element Stokes parameters can be derived from dual-polarised data such as Sentinel-1 SLC products, the challenge imposed on this usage is the separability between single and double bounce targets due to the ellipticity angle from dual-pol data on the verge of zero (Raney, 2007), which could decrease the difference between polarimetric scattering types. Despite this challenge, one scattering type between the mix of all scattering mechanisms often dominates agricultural targets at a certain growth stage, and the scattering type of dominance could change with the development of the canopy (McNairn et al., 2009). Therefore, all scattering types of m-chi decomposition are considered in this study as input predictors to classify crops across full growth stages.

4.4.4 Model implementation

Considering the model inputs, we adopted the typical remote sensing scene classification approach and extracted square image patches centred around a labelled crop category. Square patches for each crop class are generated with the size of 9×9, 11×11, 15×15 and

21×21. Larger image patches will cover multiple polygons from different crop species but could lead to greater model parameters. Therefore, the optimal size of image patches is selected to 11×11 after examinations for classification purposes. Therefore, the complete shape of input image patches for the proposed model is $11 \times 11 \times 26 \times 3$ (width × height × sequence × channel). Patches without any ground truth information and missing values of pixels were eliminated.

The implementations of models are customised to train the network with reference to the Adam optimiser (Kingma and Ba, 2014), and cross-entropy loss function (Botev et al., 2013). The learning rate for the optimiser is set to 0.001, and weight decay is regularised at 0.0001. A batch size of 128 is used during the training stage, and the model is saved with the best validation accuracy for later model inference. Another regularisation strategy, namely Early stopping, is used to prevent the models from overfitting since it will terminate the model training progress once the model validation performance is relatively stabilised during the iterative training process.

The proposed 3D-ConvSTAR is compared with deep learning architectures applied in other studies: a 1D-CNN-based architecture, namely temporal CNN (TCNN) (Pelletier et al., 2019), a typical 3D-CNN (Ji et al., 2018), a hybrid 3D-2D CNN (Roy et al., 2019) and a 3-layer ConvSTAR (Turkoglu et al., 2021b). This study reproduced the implementations of the models in the aforementioned studies. The modelling environment is implemented in Python 3.7.15 with Tensorflow backend (2.5.0) and Keras library (2.1.1) for model construction and generalisation under two graphic devices of NVIDIA Quadro P4000 (8 GB RAM per GPU), and two processors of Intel (R) Xeon (R) Silver 4114 CPU (2.20GHz/2.19 GHz).

4.4.5 Data augmentation

Imbalanced class distribution in real-world label datasets often leads to bias in supervised classification approaches, where machine learning models, under typical model training schemes, are prone to weigh importance in favour of majority classes (Ren et al., 2018; Dong et al., 2018). Underrepresented crop classes in certain places may have the same or even higher value (either financial or ecological) as the crops that occur more frequently (Turkoglu et al., 2021a). The annual statistics record that the total sown area for wheat is 6,309 hectares compared to 64,564 and 135,401 hectares for maize and soybean in Bei'an 2017 (Heihe Social

and Economic Statistics Yearbook, 2018). Consequently, class-balanced model evaluation is crucial for agricultural mapping applications, as fine-structured agricultural systems in local areas could lead to crop classes with the imbalanced distribution.

One approach to address this issue is typically called inverse frequency weighting (Cui et al., 2019). The training samples of minority classes are assigned higher weighting factors than majority classes for calculating training loss, which neutralizes the model bias towards majority classes. The underlying side effect of weighting samples could weaken model performance on majority classes to which lower weights are given. Another approach to balance the dataset is either via oversampling minority classes (Ling and Sheng, 2008) or undersampling majority classes (He and Garcia, 2009). Both techniques depend on a trade-off for the number of samples across all classes. Missing data for dominant classes due to undersampling could severely affect model performance, considering that deep learning and machine learning methods are data-driven and data-hungry. Therefore, undersampling is not used in this study.

This study applied an oversampling technique followed by a rotation of image patches. Each resampled sample patch is randomly rotated within 180 degrees and flipped horizontally. Oversampling minor classes, however, is not always a panacea, as it might not significantly improve mapping minority classes when the sample size is too small. A recent data augmentation method called mix-up, designed to linearly combine labelled images for model training, has proven successful in mapping tree species (Mäyrä et al., 2021). In this study, the input image patches of different crop classes and corresponding labels (categorical encoding) are mixed up, respectively, to generate blended or synthetic datasets, i.e., an image patch may contain 20% soybean and 80% wheat. The class distribution of the output via mix-up is also balanced. This study proposes a joint learning structure to combine deep learning models to counter imbalanced class distribution and compares it with the aforementioned data augmentation techniques, including oversampling, inverse weighting, and mix-up, for mapping minority classes.

4.4.6 Model interpretation

The interpretability of deep learning models on crop mapping tasks is still limited, considering that the extracted higher-level features are outputted by a hidden learning process in the operating mechanism held by deep learning approaches, which is often called a ‘black box’

(Heo et al., 2019). The explanation of deep learning methods benefits users in understanding the intricate patterns of crop growth and evaluating model reliability on crop mapping (Zhong et a., 2019; Xu et al., 2021). Previous studies investigated deep learning methods by visualizing intermediate layers of the networks for monitoring the model learning process and temporal learning patterns of certain crops (Xu et al., 2020; Zhong et al., 2019). Another approach to interpreting deep learning models is based on gradient-based explanations (Xu et al., 2021; Bastings and Filippova, 2020; Rußwurm and Körner., 2020; Mäyrä et al., 2021). Typically, the deep learning model input and corresponding labels are fitted with neural-network-based functions that are differentiable and perform nonlinear transformations. According to the gradient descent algorithm, the model weights during feature extraction can be iteratively updated and optimized to minimize the difference between predicted outputs and corresponding true input values. This study computed the gradients of the predicted scores for each crop type with respect to input image patches for the proposed model via vanilla backpropagation. A gradient for each crop is composed of an array of partial derivatives and it signifies the correlation between the changes in the input features and the corresponding prediction score. The highest magnitude of the gradient indicates the most influential pixels for the process of identifying certain crops. The prediction score is the model soft output derived by the softmax function at the last layer of the proposed model (Figure 4-5) and it suggests the confidence degree of the proposed model to the classification results of each crop category in this study. The gradients for each class can be visualised via saliency maps. Considering that the spatial dimension of the sample patches used in this study is only 11×11, the assessment of the important input features at spatial dimension may not decisively and accurately demonstrate the locations that contribute to spatial importance for crop mapping. Therefore, the results are more of a performance check for the proposed model.

4.4.7 Evaluation metrics

For the accuracy assessment of each network, overall accuracy (OA), Cohen's kappa coefficient (Kappa), and F1 score were selected as the performance indicators in this study. The OA is calculated to evaluate the overall model performance. Overall accuracy is calculated by aggregating the number of correctly classified values $n_i^{correct}$ based on the number of classes C and dividing by the total number of samples N in Eq. (4-10):

$$OA = \frac{\sum_{i=1}^{C} n_i^{correct}}{N} \tag{4-10}$$

The Kappa coefficient, also known as Cohen's Kappa, is a widely used metric in deep learning and remote sensing studies to assess the performance of classification models, particularly in the context of crop classification (Congalton, 1991; Foody, 2004). It is computed from the empirical probability of observed agreement, also known as OA and expected agreement p_e in Eq. (11) and (12). p_e is calculated by n_i^p, the total number of predicted labels, and n_i^t, the total number of ground truth labels. Kappa values range from -1 to 1, with values closer to 1 interpreted as a high level of agreement between the predicted and ground truth labels and values closer to 0 indicating that the agreement is no better than chance. In remote sensing and crop classification studies, a Kappa coefficient is often used alongside other performance metrics, such as the overall accuracy, producer's accuracy, and user's accuracy, to provide a comprehensive evaluation of the classification model (Congalton, 1991):

$$Kappa = \frac{OA - p_e}{1 - p_e} \tag{4-11}$$

$$p_e = \frac{\sum_{i=1}^{C} n_i^p n_i^t}{N^2} \tag{4-12}$$

The F1 score is used to measure classification performance grouped into categories since the sample data were imbalanced in this study. It relates to the harmonic mean of the producer's accuracy ($Recall$) and user's accuracy ($Precision$) respectively (Stehman., 2001) and is determined as follows:

$$F1_i = 2 \cdot \frac{Precision_i \cdot Recall_i}{Precision_i + Recall_i}, i \in C \tag{4-13}$$

4.5 Results

4.5.1 Classification results

It is found that the deep learning models using m-chi decomposition features yielded better classification results than using backscatter and its cross-ratio. See Table 4-1. Especially for TCNN, this 1D-CNN-based architecture benefited from polarimetric features significantly, increasing the OA and Kappa by > 20% and > 30%, respectively. With regard to the models considering both spatial and temporal dimensions, using m-chi decomposition features slightly improved the classification accuracy over backscatter, but these models outperform TCNN. Compared with other models, the proposed method, 3D-ConvSTAR, achieved the highest OA

on both backscatter (82.0%) and m-chi decomposition (89.4%). On the other perspective, incorporating few optical acquisitions into the sequential radar dataset contributes to improved performance for all models compared to using standalone multi-temporal SAR data, and the deep learning models with the combination of multispectral bands and polarimetric features performed the best among all scenarios. Under this circumstance, the proposed method, 3D-ConvSTAR, outperformed other approaches in terms of the highest OA (91.7%) and Kappa (85.7%). The second-best performer is the standalone 3D-CNN under the same scenario.

Table 4-1. The comparison of model performance based on multiple composite features. The best scores for each metric are highlighted in bold, and the second best are underlined.

Features \ Models		TCNN	3D-CNN	3D-2D CNN	ConvSTAR	3D-ConvSTAR
Backscatter	OA (%)	56.7	82.3	79.2	81.9	82.0
	Kappa (%)	20.9	68.6	62.9	68.0	68.2
m-chi	OA (%)	77.2	85.0	84.1	85.8	89.4
	Kappa (%)	59.4	73.6	71.8	75.2	81.6
Optical+backscatter	OA (%)	72.6	87.0	87.4	88.9	87.9
	Kappa (%)	51.0	77.5	78.1	80.5	79.0
Optical+m-chi	OA (%)	86.4	<u>90.5</u>	89.0	90.0	**91.7**
	Kappa (%)	75.8	<u>83.5</u>	81.1	82.6	**85.7**

Table 4-2 presents the F1-score yielded by deep learning models for each crop type, based on different combinations of input features. The polarimetric parameters overall are better than backscatter according to the F1 scores derived by each model for crop types. Especially m-chi decomposition features lead to significant improvement for TCNN on differentiating crops compared to using backscatter. For the models that consider the spatio-temporal dimension, m-chi decomposition features are still reliable predictors over backscatter for identifying maize and soybean, and 3D-ConvSTAR its advantage on the minority class by producing 81.8% of F1 for wheat that is the highest accuracy among the model results with connection to applying m-chi decomposition. For the use of a combination of multispectral bands and SAR features, ConvSTAR produced the highest F1 scores for maize (93.8%) and soybean (92.3%), which are only 0.1% higher than the same measurements derived by 3D-ConvSTAR. However, 3D-ConvSTAR also yielded the leading performance on less frequent classes, including wheat (90.9%) and other crops (74.0%), resulting in the highest mean F1 (87.7%) among all types of input features used. Therefore, the best-performing model for crop classification with an imbalanced dataset is the proposed 3D-ConvSTAR with the combined features of m-chi decomposition and multispectral bands.

Table 4-2. The comparison of model performance in each class. The best score for each column is highlighted with bold and the second best is underlined.

Models	Features	Maize F1 (%)	Soybean F1 (%)	Wheat F1 (%)	Other crops F1 (%)	Mean F1 (%)
TCNN (Pelletier et al., 2019)	Backscatter	62.2	56.2	2.5	0.6	30.4
	m-chi	80.9	79.0	37.6	38.2	58.9
	Optical+backscatter	78.1	71.2	63.2	1.1	53.4
	Optical+m-chi	91.6	88.1	43.1	36.2	64.8
3D-CNN (Ji et al., 2018)	Backscatter	86.1	84.5	50.4	41.2	65.6
	m-chi	89.0	86.5	60.8	47.7	71
	Optical+backscatter	90.6	88.4	75.8	50.4	76.3
	Optical+m-chi	93.5	91.3	69.0	<u>70.0</u>	81.0
3D-2D CNN (Roy et al., 2019)	Backscatter	82.5	81.9	6.5	45.5	54.1
	m-chi	87.8	86.3	12.1	54.1	60.1
	Optical+backscatter	91.1	89.5	68.1	48.7	74.4
	Optical+m-chi	93.4	91.1	77.6	62.2	81.1
ConvSTAR (Turkoglu et al., 2021b)	Backscatter	85.5	84.4	30.8	45.9	61.7
	m-chi	90.0	88.3	22.2	55.5	64.0
	Optical+backscatter	93.5	90.0	76.2	38.1	74.5
	Optical+m-chi	**93.8**	**92.3**	79.7	52.5	79.6
3D-ConvSTAR	Backscatter	85.6	83.6	69.4	38.0	69.2
	m-chi	92.1	90.4	81.8	61.0	<u>81.3</u>
	Optical+backscatter	91.7	89.6	<u>83.9</u>	45.6	77.7
	Optical+m-chi	<u>93.7</u>	<u>92.2</u>	**90.9**	**74.0**	**87.7**

The classification performance using 'Optical+m-chi' as input features varies across data augmentation techniques, with some models performing better or worse than the baseline where no data augmentation is applied in the training data. (Table 4-3). The oversampling method generally performed better than other augmentation methods. TCNN has been improved with oversampling significantly from 64.8% to 83.5% on mean F1. In contrast, balanced loss and mix-up reduced the overall performance for 3D-2D CNN and 3D-ConvSTAR. Some data augmentation methods for the certain model, such as 3D-ConvSTAR, marginally increased identification performance for majority crops including maize and soybean by 0.5% and 0.1% respectively while lowered the performance for minority classes. Furthermore, each method is compared based on model predictions for different geographical locations within Bei'an, taking into account the highest mean F1 score achieved by certain models that incorporate data augmentation techniques (Figure 4-6).

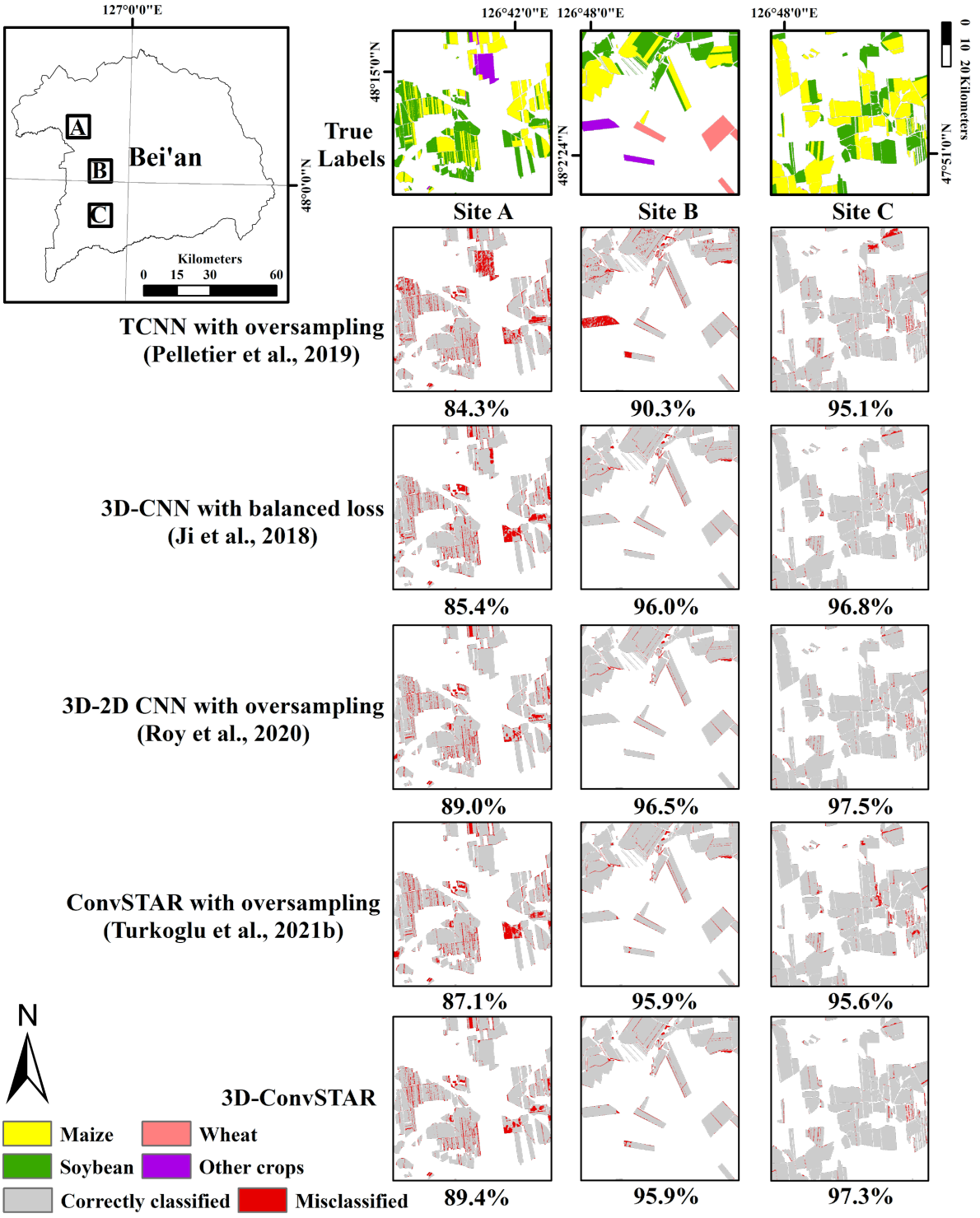

Figure 4-6. Comparison of classification performances between the models with applying data augmentation techniques and the proposed method across various sites within Bei'an. Percentages indicate the proportion of correctly classified samples with respect to ground truth labels.

Table 4-3. The comparison of model performance with applying data augmentation techniques. The best measurements for each column are highlighted in bold, followed by underlines indicating the second-best performance.

Models	Inputs	Methods	Maize F1 (%)	Soybean F1 (%)	Wheat F1 (%)	Other crops F1 (%)	Mean F1 (%)
TCNN (Pelletier et al., 2019)	Optical+m-chi	Oversampling	92.7	90.7	83.9	66.8	83.5
		Balanced loss	91.5	88.7	64.1	49.1	73.4
		Mix-up	91.6	89.9	63.3	50.0	73.7
			91.6	88.1	43.1	36.2	64.8
3D-CNN (Ji et al., 2018)	Optical+m-chi	Oversampling	92.6	90.2	88.4	59.7	82.7
		Balanced loss	<u>94.0</u>	<u>92.2</u>	86.1	71.1	85.9
		Mix-up	93.9	91.9	66.3	65.5	79.4
			93.5	91.3	69.0	70.0	81.0
3D-2D CNN (Roy et al., 2019)	Optical+m-chi	Oversampling	92.2	90.3	87.1	60.6	82.6
		Balanced loss	92.7	88.2	74.6	53.4	77.2
		Mix-up	91.9	90.4	62.2	62.0	76.6
			93.4	91.1	77.6	62.2	81.1
ConvSTAR (Turkoglu et al., 2021b)	Optical+m-chi	Oversampling	92.8	90.7	**91.8**	71.2	<u>86.6</u>
		Balanced loss	93.0	90.4	85.3	55.4	81.0
		Mix-up	**94.2**	91.4	85.8	<u>71.5</u>	85.7
			93.8	**92.3**	79.7	52.5	79.6
3D-ConvSTAR	Optical+m-chi	Oversampling	93.8	91.2	85.6	67.6	84.6
		Balanced loss	93.6	90.7	89.5	63.6	84.4
		Mix-up	**94.2**	**92.3**	88.5	71.0	86.5
			93.7	<u>92.2</u>	<u>90.9</u>	**74.0**	**87.7**

A close examination of the classification accuracies reveals that the proposed 3D-ConvSTAR method outperforms the other models in most cases. At Site A, the 3D-ConvSTAR model yields an accuracy of 89.4%, which is higher than the accuracies derived by the TCNN with oversampling (84.3%), 3D-CNN with balanced loss (85.4%), the 3D-2D CNN with oversampling (89.0%) and ConvSTAR with oversampling (87.1%) models. At Site B, the 3D-ConvSTAR method and ConvSTAR with oversampling both achieve a classification accuracy of 95.9%. While the 3D-2D CNN with oversampling model exhibits a slightly higher accuracy (96.5%), the 3D-ConvSTAR model outperforms the other two methods, namely TCNN with oversampling (90.3%) and 3D-CNN with balanced loss (96.0%). Finally, at Site C, the proposed 3D-ConvSTAR model demonstrates the second-highest classification accuracy (97.3%), only surpassed by the 3D-2D CNN with oversampling (97.5%). Nevertheless, the difference is marginal and 3D-ConvSTAR method outperforms the rest of the scenarios. In summary, the proposed 3D-ConvSTAR demonstrates competitive performance in comparison to the other deep learning models fed with augmented data for crop mapping across all three sites. It also performs the highest mean F1 score of 87.7% on the testing dataset compared with

the models using augmented data (Table 4-3). Therefore, the proposed method was used to predict unseen data for the creation of a thematic map.

As mentioned in Figure 4-2, ten per cent of total ground truth samples were used for dataset split and sixty percent of which was used for model training. Then, the pre-trained 3D-ConvSTAR produced the annual crop map during model inference for Bei'an in 2017, shown in Figure 4-7. The predicted results were also compared with the total ground truth dataset in the confusion matrix illustrated in Figure 4-8. The cell in the top left corner shows that 96.58% of maize instances were correctly classified as maize, while 2.92% of them were misclassified as soybean and 0.00% as wheat. Similarly, the cell in the bottom right corner shows that 86.87% of instances for other crops were correctly classified as themselves, while 11.18% of them were misclassified as soybean. In general, the model performs relatively well in distinguishing maize, soybean, and wheat while it struggles more with identifying other crops.

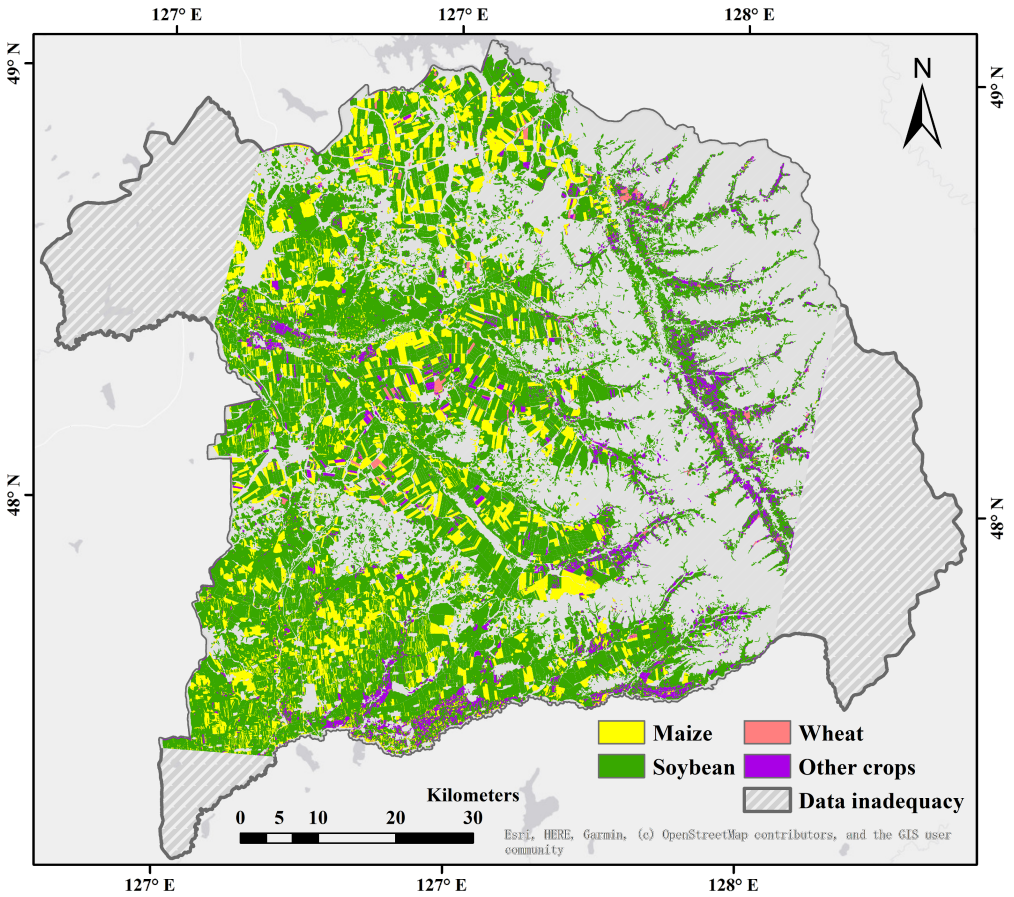

Figure 4-7. The annual crop map for Bei'an 2017. It was produced by 3D-ConvSTAR, weakly supervised with ten per cent of all ground truth samples. The areas not designated as cropland were excluded using a cropland mask introduced in Section 3.2. Data inadequacy indicates the absence of data collected from Sentinel-1 and Sentinel-2 images not fully covering the study area throughout the crop growth season, with areas identified outside the overlapping area in Figure 4-1, suggesting incomplete imaging and insufficient temporal coverage.

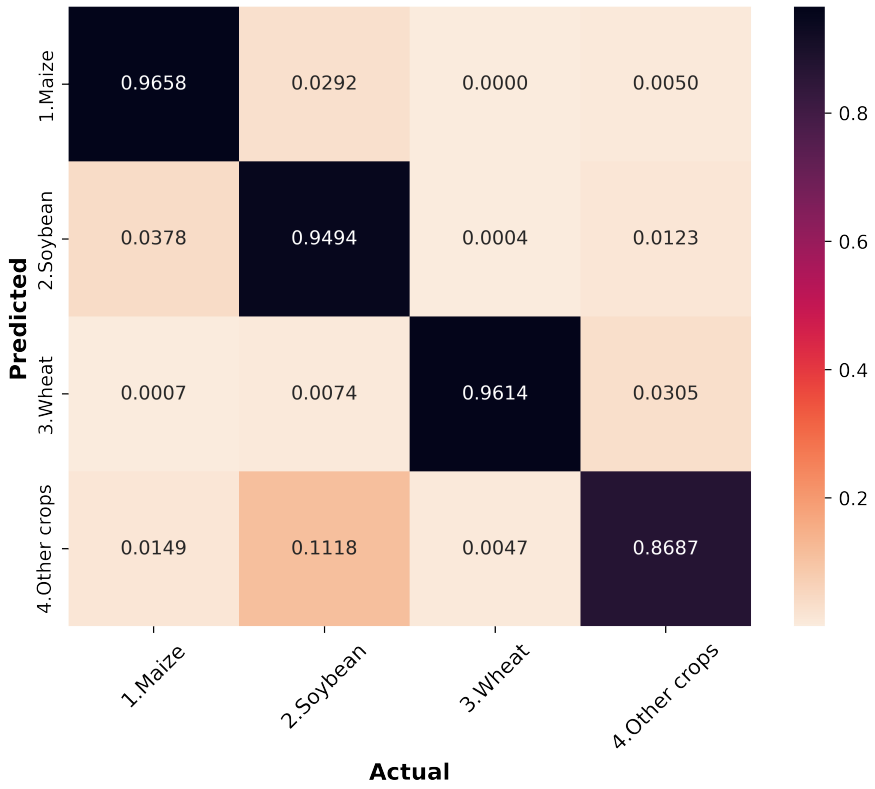

Figure 4-8. The confusion matrix for the comparison between predicted labels derived by 3D-ConvSTAR and all ground truth labels.

4.5.2 Model interpretation

The prediction scores, as outlined in Section 4.4.6, are soft outputs produced by the final layer of the 3D-ConvSTAR model (illustrated in Figure 4-5). Figure 4-9 visually demonstrates the confidence level of the proposed model in its crop classification performance on the testing dataset. The results indicate that 3D-ConvSTAR exhibits a higher level of confidence in its mapping of maize, soybean, and wheat crops, as compared to other crops. This is evident from the concentration of prediction scores for most samples of the three classes, which averagely hover around 90%. The proposed model is less confident in accurately identifying other crops with reference to the relatively lower mean prediction score (71%), which is consistent with the misclassification presented by the F1 score in Table 4-2, despite the fact that the 'other crops' category has a larger number of training samples than wheat (See Figure 4-2).

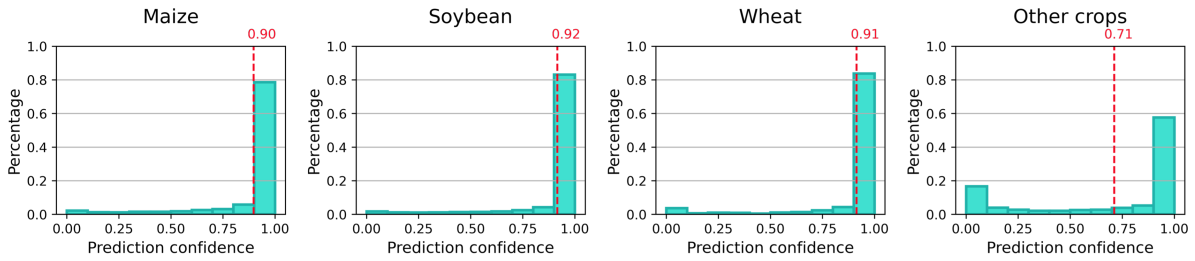

Figure 4-9. The prediction score distribution for each crop, derived by the last dense layer of 3D-ConvSTAR. The red dashed lines indicate average prediction scores.

The average gradient magnitudes are represented by saliency maps, as shown in Figure 4-10. These maps represent the most important spatial locations for each crop type in the image samples of the testing dataset. The shape of the pixel chunks with the highest importance for maize may complement the part of the lowest importance for soybean and vice versa, indicating that maize and soybean in the samples are mostly intercropped. The pixel importance for wheat shows that the field shapes in the samples are mostly separated. However, the pixel importance for other crops is scattered without forming a clear shape, which corresponds to the relatively lower mapping accuracy for those crops.

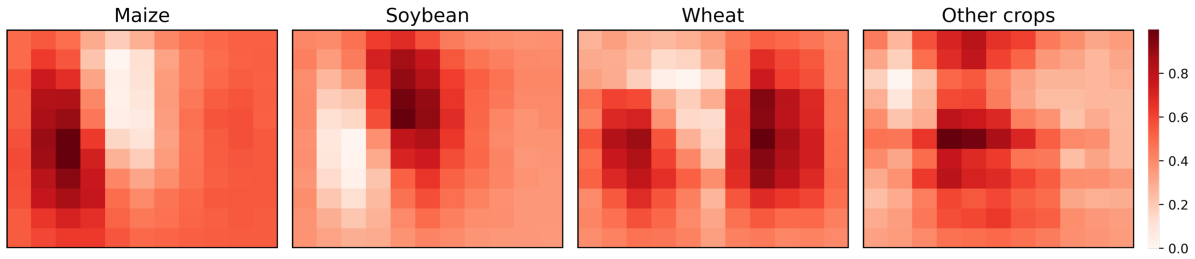

Figure 4-10. The saliency maps represented by the average magnitude of gradients for each crop. 1500 image patches were randomly extracted from the testing dataset and fed into 3D-ConvSTAR to generate saliency maps for illustration.

In summary, the 3D-ConvSTAR model shows high confidence in classifying maize, soybean, and wheat, with prediction scores for these crops averaging around 90%, as evidenced in testing. In contrast, the model is less confident in classifying other crop types, demonstrated by a lower average prediction score of 71%. Saliency maps reveal that the model identifies distinct spatial features for each crop type. Maize and soybean often appear intercropped, as indicated by complementary areas of importance in their respective saliency maps. Wheat, however, is typically mapped in separate fields, reflected by dispersed spatial patterns in the imagery. For other crops, the saliency maps show scattered and undefined patterns aligned with their lower classification accuracy.

4.6 Discussion

In this study, various deep learning architectures for patch-based crop mapping are evaluated using SAR and fused SAR-optical data. The proposed method functions as an ensemble deep learning by synergistically connecting 3D-CNN and ConvSTAR, reaching the highest performance overall among all models in terms of the OA (91.7%) and Kappa (85.7%) for the identification and classification of different crops (see Table 4-1). This study also validated the effectiveness of using SAR polarimetric decomposition parameters of m-chi for identifying certain crops over using SAR backscatter, which confirmed the findings by previous studies (e.g., De et al., 2014; Sonobe et al., 2019). The integration of a few optical acquisitions and time-series SAR image data does improve the classification performance overall compared to standalone SAR data, but maize and soybean are not increased significantly (< 10% on average) regarding the F1-score derived by all models (Table 4-2). These results are likely to be explained by the fact that maize and soybean are dominant crops in the county that provides an enormous number of labelled ground truth (Figure 4-2), so both crops can be well-trained by those data-driven models. In contrast, the fusion of optical and m-chi decomposition features enhanced minority classes surprisingly by all models, especially for wheat. The proposed method with multisource fusion yields the highest F1-score (90.9%) for wheat compared to using backscatter (69.4%) and polarimetric features (81.8%). This indicates that multispectral information contributes mostly to the enhancement of mapping for minority crops.

We discovered that crop mapping performance can be enhanced not only by fusing SAR-optical datasets, which provide spatio-temporal, polarimetric, and spectral characteristics related to different crop structures (Gao et al., 2018; Van Tricht et al., 2018), but also by utilizing multispectral information as a reliable complementary source owing to the synergistic nature of SAR and optical data. This study demonstrates that Sentinel-1 and Sentinel-2 imagery exhibit a mutually complementary effect, increasing the sensitivity of both sensors to specific crop class characteristics throughout the growing season. Sentinel-2 data can be associated with the quantitative analysis of chlorophyll and moisture content in crop leaves, with spectral bands such as the Vegetation Red Edge being particularly useful for differentiating certain crops, confirming previous findings by Guerschman et al. (2003) and You et al. (2020). Sentinel-1 data is sensitive to morphological variations, as it provides biophysical, structural, and agronomic characteristics, and is strongly correlated with the structural development of crops during the growing season (Adrian et al., 2021; Sonobe, 2019).

In addition to the SAR backscatter, polarimetric parameters can reflect in-depth scattering properties of crops due to scattering mechanisms with robust physical interpretability, making them useful for crop mapping (He et al., 2020; Liao et al., 2020; Xie et al., 2019; Gao et al., 2018). Polarimetric features directly relate to the underlying physical properties of the crops. These parameters can be used to quantify the contribution of different scattering mechanisms and provide insights into the crops' biophysical, structural, and agronomic properties, such as crop type, plant density, leaf area index, and growth stage. In this study, we capitalised SAR scattering diversity, such as surface scattering, volume scattering, and double-bounce scattering (See Section 4.4.3) that are responsible for the interactions between the electromagnetic waves emitted by SAR sensors and the various components of a crop's structure. Sentinel-1 data, however, is limited to insufficient decomposition methods, since it is a dual-polarised SAR sensor, which restricts the available decomposition methods. While quad-polarimetric data can be analysed using fully polarimetric decomposition algorithms, this study found that m-chi decomposition, originating from compact-pol planforms, effectively maps crops using dual-polarisation data and synergizes well with optical bands. It is important to note that fully polarimetric SAR systems often have reduced swath coverage and relatively inconsistent temporal frequency, posing challenges for crop mapping across extensive areas (Sonobe et al., 2019).

With respect to the ablation study regarding applying different input features, the proposed deep learning network showcases the advantage of the proposed model that outperformed the deep learning methods with standalone architectures (TCNN, 3D-CNN and ConvSTAR), and combined architectures such as 3D-2D CNN proposed in other studies for crop mapping under the same input predictors, as detailed in Table 4-2. The model performance varies significantly across different crop classes, with wheat and other crops generally exhibiting lower F1 scores compared to maize and soybean. The 3D-ConvSTAR improved all performance overall, in particular per-class performance in separating the maize, soybean and wheat, providing a beneficial method for local industries due to the commercial interests in these crops. We also investigated the comparative analysis of various deep learning models with data augmentation techniques for crop mapping to deal with imbalanced class distribution (Table 4-3). Examining the F1 scores for individual crop classes, it is evident that the 3D-ConvSTAR model using the mix-up consistently outperforms other combinations, achieving 94.2% for maize, and 92.3% for soybean, but reducing performance for wheat and other crops. All models produced similar results for maize and soybean after data augmentation techniques are applied, and

oversampling generally outperforms other augmentation methods. Wheat and other crops generally exhibit lower F1 scores compared to maize and soybean. This finding, while preliminary, may imply that this data augmentation technique could be more useful for enriching training samples when data collection is a major challenge in certain research fields. For example, it is particularly well-suited to augment imagery data collected from the human nervous system (Smucny et al., 2022), and increase the airborne training sample size for mapping species (Mäyrä et al., 2021). However, it may not be useful to overcome imbalanced class distribution.

The proposed network, in terms of average F1 score, outperforms other approaches that rely on data augmentation techniques, primarily because it effectively integrates the temporal nature of remote sensing data into a more sophisticated input space while accounting for the spatial relationships between features along the time series. This leads to a better separation of crop types with homogeneous representations (Figure 4-7 and Figure 4-8). However, the proposed method is prone to generating a higher number of training parameters compared to alternative methods, resulting in increased model training time. This issue is particularly due to the connections between the learned features produced by the 3D-CNN, ConvSTAR, and the subsequent shallow CNNs implemented by the 2D-CNN, as these settings can lead to an increased number of training parameters. All classifiers exhibit suboptimal performance for the 'other crops' category, which can be attributed to the mixture of various crop types. Each of these unknown crop types is only represented by a relatively small sample size in the training data, thereby limiting the model's ability for identification. Although the "other crops" category has more training samples than wheat (Figure 4-2), it may still be underrepresented compared to samples for maize, soybean and wheat in the dataset. This could lead the model to focus on the distinctive class labels and consequently perform poorly in the "other crops" category with mixed labels for unknown crop types. The "other crops" category may encompass a wide range of crop types, each with distinct spectral, polarimetric and temporal signatures. For example, there is only one field parcel for rice in the ground truth data collection, so it was labelled as other crops in this study. This increased diversity may make it more challenging for the deep learning model to accurately identify and classify these crops. In contrast, maize, soybean, and wheat may have more consistent and easily distinguishable characteristics, allowing for a higher F1 score. Additionally, the features extracted from the combination of SAR and optical data might be more informative and discriminative for maize, soybean, and wheat than for the "other crops" category. The complexity of the features for the "other crops" category might be

higher, making it more difficult for the model to learn and correctly classify these samples, which leads to the lower F1 score observed.

The results overall demonstrate the advantages of using the combination of polarimetric and multispectral data from Sentinel-1 and Sentinel-2. Data fusion provides additional information for classification algorithms to exploit crops' structural details, while also offering supplementary polarimetric and spectral properties. The discrepancy in F1 scores between the wheat and "other crops" categories in crop mapping highlight the need for further research into optimizing model architectures and methodologies to enhance crop mapping across all classes. Moreover, future work should focus on integrating fully polarimetric data with optical data to further improve crop mapping accuracy by applying popular deep learning architectures, such as Fully Convolutional Neural Networks (FCN), which perform pixel-wise segmentation on images. For instance, the 3D U-Net architecture can be employed to extract the spatio-temporal features of crops (Ji et al., 2018; Adrian et al., 2021). Investigating the contribution of SAR texture information combined with optical data to semantic segmentation for crop mapping also enables further exploration. Recent studies have employed a self-attention-based convolutional recurrent network to learn temporal dependencies of multivariable time series (Fu et al., 2022) and combined 3D-CNN with an attention-based recurrent network for crop yield prediction (Nejad et al., 2022). Both studies assessed the feasibility of attention mechanisms in extracting attentive spatio-temporal features. Consequently, future research could involve the integration of 3D-CNN with attention-based convolutional recurrent networks, such as ConvSTAR, for crop mapping and comparison with architectures for semantic segmentation. More importantly, the model's robustness should be further evaluated for predicting crop types in different years. Model behaviours may be influenced by interannual variability within the same region, and recurrent structures have shown promise in capturing crop phenological characteristics and enhancing model generalisation (Xu et al., 2021). Assessing the model's spatial transferability is also a critical aspect of future research, given the potential application of the model in diverse geographical contexts. This could facilitate the design of efficient strategies for improved applicability, potentially contributing to the optimisation of agricultural practices and crop mapping on a global scale. One such strategy refers to training the model with representative crop datasets that can accurately reflect the complexity and heterogeneity of the agricultural landscape. Alternatively, the model's parameters could be adjusted to accommodate the unique conditions of specific locations.

4.7 Conclusion

In this research, we proposed a workflow for multi-temporal crop mapping based on the fusion of Sentinel-1 polarimetric parameters and Sentinel-2 multispectral reflectance, combined with various deep learning architectures. The proposed 3D-ConvSTAR, which connects 3D-CNN layers and convolutional recurrent layers, delivers enhanced classification performance for crop mapping in comparison to the architectures designed in previous studies. Additionally, the designed architecture is robust when training the dataset with imbalanced class distribution and outperforms other data augmentation techniques. This study demonstrates that crop mapping can be conducted with high accuracy using the proposed 3D-ConvSTAR in terms of overall accuracy and F1 score for each crop class. Although the implemented architecture is likely not the optimal solution, given the training parameters overload, it still manages to produce accurate and valuable results for separating the crops with significant commercial value in Bei'an. While the proposed network exhibits superior performance in terms of crop type separation and accounting for the temporal and spatial relationships in remote sensing data, it is essential to address the challenges posed by the increased number of training parameters and the inherent limitations in classifying underrepresented crop types. Future research should focus on optimizing the network architecture and exploring alternative approaches to improve classification accuracy across all crop types while minimizing the computational cost associated with training the model. The model's generalisation for crop mapping needs further assessment based on interannual and spatial transferability.

References

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

Ainsworth, T.L., Kelly, J.P. and Lee, J.S., 2009. Classification comparisons between dual-pol, compact polarimetric and quad-pol SAR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(5), pp.464-471.

Audebert, N., Le Saux, B. and Lefèvre, S., 2019. Deep learning for classification of hyperspectral data: A comparative review. *IEEE Geoscience and Remote Sensing Magazine*, *7(2)*, pp.159-173.

Bargiel, D., 2017. A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sensing of Environment*, *198*, pp.369-383.

Bastings, J. and Filippova, K., 2020. The elephant in the interpretability room: Why use attention as explanation when we have saliency methods?. *arXiv preprint arXiv:2010.05607*.

Boryan, C., Yang, Z., Mueller, R. and Craig, M., 2011. Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto International*, *26*(5), pp.341-358.

Botev, Z.I., Kroese, D.P., Rubinstein, R.Y. and L'Ecuyer, P., 2013. The cross-entropy method for optimization. In *Handbook of Statistics*, *Vol. 31*, pp. 35-59.

Boulila, W., Ghandorh, H., Khan, M.A., Ahmed, F. and Ahmad, J., 2021. A novel CNN-LSTM-based approach to predict urban expansion. *Ecological Informatics*, *64*, p.101325.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*(1), pp.35-46.

Crisóstomo de Castro Filho, H., Abílio de Carvalho Júnior, O., Ferreira de Carvalho, O.L., Pozzobon de Bem, P., dos Santos de Moura, R., Olino de Albuquerque, A., Rosa Silva, C., Guimaraes Ferreira, P.H., Fontes Guimarães, R. and Trancoso Gomes, R.A., 2020. Rice crop detection using LSTM, Bi-LSTM, and machine learning models from Sentinel-1 time series. *Remote Sensing*, *12*(16), p.2655.

Cui, Y., Jia, M., Lin, T.Y., Song, Y. and Belongie, S., 2019. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9268-9277.

De, S., Kumar, V. and Rao, Y.S., 2014, June. Crop classification using RISAT-1 hybrid polarimetric SAR data. In *EUSAR 2014; 10th European Conference on Synthetic Aperture Radar*, pp. 1-4.

Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaute, T. and Savinaud, M., 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*, pp.551-568.

Dong, J., Xiao, X., Zhang, G., Menarguez, M.A., Choi, C.Y., Qin, Y., Luo, P., Zhang, Y. and Moore, B., 2016. Northward expansion of paddy rice in northeastern Asia during 2000–2014. *Geophysical Research Letters*, *43*(8), pp.3754-3761.

Dong, Q., Gong, S. and Zhu, X., 2018. Imbalanced deep learning by minority class incremental rectification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(6), pp.1367-1381.

Dou, P., Shen, H., Li, Z. and Guan, X., 2021. Time series remote sensing image classification framework using combination of deep learning and multiple classifiers system. *International Journal of Applied Earth Observation and Geoinformation*, *103*, p.102477.

Fan, C., Zheng, B., Myint, S.W. and Aggarwal, R., 2014. Characterizing changes in cropping patterns using sequential Landsat imagery: An adaptive threshold approach and application to Phoenix, Arizona. *International Journal of Remote Sensing*, *35*(20), pp.7263-7278.

Foody, G.M., 2004. Thematic map comparison. *Photogrammetric Engineering & Remote Sensing*, *70*(5), pp.627-633.

Fu, E., Zhang, Y., Yang, F. and Wang, S., 2022. Temporal self-attention-based Conv-LSTM network for multivariate time series prediction. *Neurocomputing*, *501*, pp.162-173.

Gao, H., Wang, C., Wang, G., Zhu, J., Tang, Y., Shen, P. and Zhu, Z., 2018. A crop classification method integrating GF-3 PolSAR and Sentinel-2A optical data in the Dongting Lake Basin. *Sensors*, *18*(9), p.3139.

Griffiths, P., Nendel, C. and Hostert, P., 2019. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sensing of Environment*, *220*, pp.135-151.

Guerschman, J.P., Paruelo, J.M., Bella, C.D., Giallorenzi, M.C. and Pacin, F., 2003. Land cover classification in the Argentine Pampas using multi-temporal Landsat TM data. *International Journal of Remote Sensing*, *24*(17), pp.3381-3402.

He, C., He, B., Tu, M., Wang, Y., Qu, T., Wang, D. and Liao, M., 2020. Fully convolutional networks and a manifold graph embedding-based algorithm for polsar image classification. *Remote Sensing*, *12*(9), p.1467.

He, H. and Garcia, E.A., 2009. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), pp.1263-1284.

Heihe Social and Economic Statistics Yearbook. 2018. *Heihe Social and Economic Statistics Yearbook*. Beijing: China Statistical Publishing House.

Heo, J., Joo, S. and Moon, T., 2019. Fooling neural network interpretations via adversarial model manipulation. *Advances in Neural Information Processing Systems*, *32*.

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P. and Koetz, B., 2015. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing*, *7*(9), pp.12356-12379.

Ji, S., Zhang, C., Xu, A., Shi, Y. and Duan, Y., 2018. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sensing*, *10*(1), p.75.

Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kussul, N., Lavreniuk, M., Skakun, S. and Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, *14*(5), pp.778-782.

Kussul, N., Mykola, L., Shelestov, A. and Skakun, S., 2018. Crop inventory at regional scale in Ukraine: developing in season and end of season crop maps with multi-temporal optical and SAR satellite imagery. *European Journal of Remote Sensing*, *51*(1), pp.627-636.

Li, C., Chen, W., Wang, Y., Wang, Y., Ma, C., Li, Y., Li, J. and Zhai, W., 2022. Mapping winter wheat with optical and SAR images based on Google Earth Engine in Henan Province, China. *Remote Sensing*, *14*(2), p.284.

Li, H., Huang, J. and Ji, S., 2019. Bearing fault diagnosis with a feature fusion method based on an ensemble convolutional neural network and deep neural network. *Sensors*, *19*(9), p.2034.

Liao, C., Wang, J., Xie, Q., Baz, A.A., Huang, X., Shang, J. and He, Y., 2020. Synergistic use of multi-temporal RADARSAT-2 and VENμS data for crop classification based on 1D convolutional neural network. *Remote Sensing*, *12*(5), p.832.

Ling, C.X. and Sheng, V.S., 2008. Cost-sensitive learning and the class imbalance problem. *Encyclopedia of Machine Learning*, *2011*, pp.231-235.

Liu, J., Kuang, W., Zhang, Z., Xu, X., Qin, Y., Ning, J., Zhou, W., Zhang, S., Li, R., Yan, C. and Wu, S., 2014. Spatiotemporal characteristics, patterns, and causes of land-use changes in China since the late 1980s. *Journal of Geographical Sciences*, *24*, pp.195-210.

Mäyrä, J., Keski-Saari, S., Kivinen, S., Tanhuanpää, T., Hurskainen, P., Kullberg, P., Poikolainen, L., Viinikka, A., Tuominen, S., Kumpula, T. and Vihervaara, P., 2021. Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sensing of Environment*, *256*, p.112322.

McNairn, H., Shang, J., Jiao, X. and Champagne, C., 2009. The contribution of ALOS PALSAR multipolarization and polarimetric data to crop classification. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(12), pp.3981-3992.

Moumni, A. and Lahrouni, A., 2021. Machine learning-based classification for crop-type mapping using the fusion of high-resolution satellite imagery in a semiarid area. *Scientifica*, *2021*.

Nair, V. and Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807-814.

Nejad, S.M.M., Abbasi-Moghadam, D., Sharifi, A., Farmonov, N., Amankulova, K. and Lászlź, M., 2022. Multispectral crop yield prediction using 3D-convolutional neural networks and attention convolutional LSTM approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, pp.254-266.

Ning, J., Liu, J., Kuang, W., Xu, X., Zhang, S., Yan, C., Li, R., Wu, S., Hu, Y., Du, G. and Chi, W., 2018. Spatiotemporal patterns and characteristics of land-use change in China during 2010–2015. *Journal of Geographical Sciences*, *28*, pp.547-562.

Nord, M.E., Ainsworth, T.L., Lee, J.S. and Stacy, N.J., 2008. Comparison of compact polarimetric synthetic aperture radar modes. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(1), pp.174-188.

Pelletier, C., Valero, S., Inglada, J., Champion, N. and Dedieu, G., 2016. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, *187*, pp.156-168.

Pelletier, C., Webb, G.I. and Petitjean, F., 2019. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, *11*(5), p.523.

Qu, Y., Zhao, W., Yuan, Z. and Chen, J., 2020. Crop mapping from sentinel-1 polarimetric time-series with a deep neural network. *Remote Sensing*, *12*(15), p.2493.

Raney, R.K., Cahill, J.T., Patterson, G.W. and Bussey, D.B.J., 2012. The m-chi decomposition of hybrid dual-polarimetric radar data with application to lunar craters. *Journal of Geophysical Research: Planets*, *117*(E12).

Ren, M., Zeng, W., Yang, B. and Urtasun, R., 2018, July. Learning to reweight examples for robust deep learning. In *International Conference on Machine Learning*, pp. 4334-4343.

Roy, S.K., Krishna, G., Dubey, S.R. and Chaudhuri, B.B., 2019. HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, *17*(2), pp.277-281.

Rußwurm, M. and Korner, M., 2017. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 11-19.

Rußwurm, M. and Körner, M., 2018. Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS International Journal of Geo-Information*, *7*(4), p.129.

Rußwurm, M. and Körner, M., 2020. Self-attention for raw optical satellite time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *169*, pp.421-435.

Sharma, A., Liu, X., Yang, X. and Shi, D., 2017. A patch-based convolutional neural network for remote sensing image classification. *Neural Networks*, *95*, pp.19-28.

Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K. and Woo, W.C., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, *28*.

Smucny, J., Shi, G., Lesh, T.A., Carter, C.S. and Davidson, I., 2022. Data augmentation with Mixup: Enhancing performance of a functional neuroimaging-based prognostic deep learning classifier in recent onset psychosis. *NeuroImage: Clinical*, *36*, p.103214.

Song, X.P., Potapov, P.V., Krylov, A., King, L., Di Bella, C.M., Hudson, A., Khan, A., Adusei, B., Stehman, S.V. and Hansen, M.C., 2017. National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey. *Remote Sensing of Environment*, *190*, pp.383-395.

Sonobe, R., 2019. Parcel-based crop classification using multi-temporal TerraSAR-X dual polarimetric data. *Remote Sensing*, *11*(10), p.1148.

Sonobe, R., Tani, H., Wang, X., Kobayashi, N. and Shimamura, H., 2014. Random forest classification of crop type using multi-temporal TerraSAR-X dual-polarimetric data. *Remote Sensing Letters*, *5*(2), pp.157-164.

Stehman, S.V., 2001. Statistical rigor and practical utility in thematic map accuracy assessment. *Photogrammetric Engineering and Remote Sensing*, *67*(6), pp.727-734.

Sun, C., Bian, Y., Zhou, T. and Pan, J., 2019. Using of multi-source and multi-temporal remote sensing data improves crop-type mapping in the subtropical agriculture region. *Sensors*, *19*(10), p.2401.

Sun, Z., Di, L., Fang, H. and Burgess, A., 2020. Deep learning classification for crop types in north dakota. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, pp.2200-2213.

Teimouri, M., Mokhtarzade, M., Baghdadi, N. and Heipke, C., 2022. Fusion of time-series optical and SAR images using 3D convolutional neural networks for crop classification. *Geocarto International*, *37*(27), pp.15143-15160.

Teluguntla, P., Thenkabail, P.S., Oliphant, A., Xiong, J., Gumma, M.K., Congalton, R.G., Yadav, K. and Huete, A., 2018. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS Journal of Photogrammetry and Remote Sensing*, *144*, pp.325-340.

Turkoglu, M.O., D'Aronco, S., Perich, G., Liebisch, F., Streit, C., Schindler, K. and Wegner, J.D., 2021a. Crop mapping from image time series: Deep learning with multi-scale label hierarchies. *Remote Sensing of Environment*, *264*, p.112603.

Turkoglu, M.O., D'Aronco, S., Wegner, J.D. and Schindler, K., 2021b. Gating revisited: Deep multi-layer RNNs that can be trained. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(8), pp.4081-4092.

Van Tricht, K., Gobin, A., Gilliams, S. and Piccard, I., 2018. Synergistic use of radar Sentinel-1 and optical Sentinel-2 imagery for crop mapping: A case study for Belgium. *Remote Sensing*, *10*(10), p.1642.

Wang, S., Azzari, G. and Lobell, D.B., 2019. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment*, *222*, pp.303-317.

Wang, T., Li, J., Zhang, M., Zhu, A., Snoussi, H. and Choi, C., 2021. An enhanced 3DCNN-ConvLSTM for spatiotemporal multimedia data analysis. *Concurrency and Computation: Practice and Experience*, *33*(2), p.e5302.

Wei, S., Zhang, H., Wang, C., Wang, Y. and Xu, L., 2019. Multi-temporal SAR data large-scale crop mapping based on U-Net model. *Remote Sensing*, *11*(1), p.68.

Xie, Q., Wang, J., Liao, C., Shang, J., Lopez-Sanchez, J.M., Fu, H. and Liu, X., 2019. On the use of Neumann decomposition for crop classification using multi-temporal RADARSAT-2 polarimetric SAR data. *Remote Sensing*, *11*(7), p.776.

Xu, J., Yang, J., Xiong, X., Li, H., Huang, J., Ting, K.C., Ying, Y. and Lin, T., 2021. Towards interpreting multi-temporal deep learning models in crop mapping. *Remote Sensing of Environment*, *264*, p.112599.

Xu, J., Zhu, Y., Zhong, R., Lin, Z., Xu, J., Jiang, H., Huang, J., Li, H. and Lin, T., 2020. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. *Remote Sensing of Environment*, *247*, p.111946.

Yang, L., Wang, L., Huang, J., Mansaray, L.R. and Mijiti, R., 2019. Monitoring policy-driven crop area adjustments in northeast China using Landsat-8 imagery. *International Journal of Applied Earth Observation and Geoinformation*, *82*, p.101892.

You, N. and Dong, J., 2020. Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, *161*, pp.109-123.

You, N., Dong, J., Huang, J., Du, G., Zhang, G., He, Y., Yang, T., Di, Y. and Xiao, X., 2021. The 10-m crop type maps in Northeast China during 2017–2019. *Scientific Data*, *8*(1), p.41.

Zhang, P. and Hu, S., 2019. Fine crop classification by remote sensing in complex planting areas based on field parcel. *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*, *35*(21), pp.125-134.

Zhao, H., Duan, S., Liu, J., Sun, L. and Reymondin, L., 2021. Evaluation of five deep learning models for crop type mapping using sentinel-2 time series images with missing information. *Remote Sensing*, *13*(14), p.2790.

Zheng, B., Myint, S.W., Thenkabail, P.S. and Aggarwal, R.M., 2015. A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *International Journal of Applied Earth Observation and Geoinformation*, *34*, pp.103-112.

Zhong, L., Gong, P. and Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. *Remote Sensing of Environment*, *140*, pp.1-13.

Zhong, L., Hu, L. and Zhou, H., 2019. Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, *221*, pp.430-443.

Zhong, L., Hu, L., Yu, L., Gong, P. and Biging, G.S., 2016a. Automated mapping of soybean and corn using phenology. *ISPRS Journal of Photogrammetry and Remote Sensing*, *119*, pp.151-164.

Zhong, L., Yu, L., Li, X., Hu, L. and Gong, P., 2016b. Rapid corn and soybean mapping in US Corn Belt and neighboring areas. *Scientific Reports*, *6*(1), p.36240.

Chapter 5 Ensemble Modelling Based on Transfer Learning for Enhancing Crop Mapping through Synergistic Integration of InSAR Coherence and Multispectral Satellite Data

Abstract

Recent advancements in remote sensing have enabled the combined use of multi-temporal and multi-modal data with sophisticated model architectures for various agricultural applications, including crop mapping. These approaches effectively and automatically correlate time series remote sensing data with crop types through multi-dimensional feature learning. This study proposes an innovative framework that explores the synergistic use of multi-temporal Sentinel-1 Interferometric Synthetic Aperture Radar (InSAR) coherence, and Sentinel-2 and RapidEye multispectral data to enhance crop mapping in smallholder croplands within Bei'an county in China. The study evaluates various deep learning models, including the 3-Dimensional U-Net (3D U-Net), Transformer, Attention-based Long Short-Term Memory (AtLSTM), and a baseline machine learning Random Forest (RF) model, focusing on their transfer learning capabilities across spatially and temporally diverse regions exhibiting complex intercropping patterns. Utilising the strengths of ensemble learning concepts, we developed a novel architecture, Transformer-AtLSTM-RF, executed under a rule-based strategy. This architecture effectively integrates different classifiers, facilitating multi-source feature fusion for enhanced crop classification performance. When fine-tuned with data specific to the the region and time period, our methodology demonstrates improved generalisability and accuracy in these complex agricultural settings. This approach yielded the highest overall accuracy (OA), mean F1 score, and mean intersection over union (mIoU) for two test sites: site A (OA: 96.2%, mean F1: 92.7%, mIoU: 86.9%) and site B (OA: 90.7%, mean F1: 88.6%, mIoU: 79.7%). Furthermore, this study evaluated input feature importance through the visualisation of dynamics of critical temporal features determined during our model inference process. This analysis interprets how different features contribute to crop identification over time, providing an in-depth understanding of underlying patterns in the feature learning process for the proposed model. Our results demonstrate the capabilities of integrated time series SAR-derived and optical data, in combination with state-of-the-art models, for mapping intercropping systems.

Keywords: crop mapping, InSAR, coherence, deep learning, transfer learning, feature importance

5.1 Introduction

Sustainable agriculture sets out goals for 'greening' growth, specifically in the context of safe and nutritious food productivity and economic viability of agricultural practices to address food security challenges in the presence of a growing global population (Cioloş and Piebalgs, 2012). The timely and precise monitoring of crop conditions, coupled with detailed spatial distribution data of croplands, is essential for agricultural sustainability, ensuring food security, developing agricultural management practices, and assessing policy decisions in the agricultural sector (Wang et al., 2013). An understanding of regional crop planting patterns predicted on accurate crop type identification is the prerequisite of strategic crop planning, especially for crops that offer rotation benefits across growing seasons, such as alternating maize and soybean rotations (Boyabatlı et al., 2019; Sahajpal et al., 2014; Wu et al., 2021; You et al., 2021). The spatial distribution of cropping patterns can reflect the land-use configurations and transitions within farmland parcels, but crop mapping performance could be impacted by the spatial heterogeneity inherent in agricultural landscapes (Zhang et al., 2021). Several studies have found that crop mapping accuracy is dependent on factors such as patch size and shape, crop planting structures, and crop density (e.g. Lechner et al., 2009; Jia et al., 2013). For instance, monocropping, characterized by single-crop farming tends to yield higher mapping accuracy compared to the mapping challenge observed in smallholder farming systems with dispersed cropland distribution (Zhang et al., 2021). In China, the complexity and difficulty of crop identification are magnified due to the agricultural heterogeneity stemming from the predominance of smallholder farming in specific local areas, especially for rotated crops with similar spectral features, growth cycles and phenological characteristics. Smallholder farmers in China typically employ crop rotation using strip cropping strategies for conservation agriculture over several decades, aiming to neutralize soil erosion and lessen dependence on mineral fertilizers (Livingston et al., 2015; Li et al., 2020). Strip cropping, a subset of the intercropping approach in which no less than two crop types are cultivated in close proximity within long and narrow multi-row strips, has been demonstrated to provide broader ecosystem services and even greater yields over single cropping in terms of enhanced spatial diversity of in-field habitats (Juventia et al., 2022). Nonetheless, there remain challenges in spatially explicit mapping of small-scale cropping patterns, which is also important for inventory considerations of crop type, location, and time.

With the availability of observations that can be repeatedly and consistently collected from multi-source satellite sensors, smallholder farmers can obtain an accurate estimation of crop dynamics across diverse spatial and temporal scales for decision-making and planning (Wen et al., 2022; Onojeghuo et al., 2023; King et al., 2017). In recent decades, optical-derived features from moderate spatial resolution platforms, such as Landsat, Sentinel-, GF-6 and MODIS, have been applied to mapping cropping intensities (Cai et al., 2018; Konduri et al., 2020; Hao et al., 2020; You et al., 2021; Zhang et al., 2022; Xia et al., 2022) and delivering annual crop maps in terms of complete crop phenological cycles within each year (Li et al., 2021; Blickensdörfer et al., 2022; Gallo et al., 2023). While optical wavelengths are intrinsically associated with crop biophysical indicators, the adaptability of microwaves to crop structural variations highlights the potential of Synthetic Aperture Radar (SAR) sensors for crop monitoring, given their all weather imaging power and penetrative capabilities into vegetation canopy targets (Veloso et al., 2017; Zhou et al., 2017; Bargiel, 2017; Steele-dunne et al., 2017; Mandal et al., 2020; Wei et al., 2021). Recent studies have explored the possibility of Sentinel-1 backscattering and its associated interferometric SAR (InSAR) coherence for comprehensive crop growth monitoring and mapping endeavours. These investigations reveal a strong correlation between different crop phenological stages within a single year and the InSAR coherence or decorrelation derived from the combination of two SAR acquisitions (Nasirzadehdizaji et al., 2021).

The integration of Sentinel-1 coherence and intensities of backscattered signals has been empirically demonstrated to enhance crop classification performance (e.g. Mestre-Quereda et al., 2020), as coherence provides information complementary to other satellite data. While optical and SAR backscatter data primarily deliver information on surface characteristics, coherence introduces additional information regarding the structural and temporal stability of vegetation fields. Coherence measurements can capture fine-scale temporal decorrelation, namely loss of coherence, across land cover, where the coherence levels decrease due to temporal decorrelation when the surface underneath changes (Sica et al., 2019), such as through various growth stages. Specifically, variations in the structure, height, and canopy coverage of different crops along their growth stages can alter the coherence signal (Blaes and Defourny, 2003). This makes coherence a dynamic indicator of changes in agricultural fields for monitoring crop growth over time and retrieving information about the imaged scene. Based on the repeat-pass interferometry and given temporal baselines, more accurate and timely agricultural mapping and crop condition monitoring can be achieved. Consequently, the combination of coherence with other satellite-derived features sets the potential to enhance the

classifier's ability to differentiate between crop types and more detailed agricultural observations. For example, the synergistic potential of SAR coherence in tandem with multispectral bands for mapping crops is less explored, although advanced crop mapping paradigms combine optical and SAR-derived features as input predictors for models (e.g. Blickensdörfer et al., 2022; Adrian et al., 2021; Bigdeli et al., 2021; Shendryk, 2019).

In parallel with advances in sensors and data availability, crop mapping has greatly benefited from the application of supervised machine learning classifiers, notably the Support Vector Machine (SVM) and Random Forest (RF). These classifiers utilize high-dimensional satellite remote sensing data and have consistently demonstrated their robustness in crop identification, (e.g. Mazarire et al., 2020; Saini and Ghosh, 2018; Bargiel, 2017; Phalke and Özdoğan, 2018; Teluguntla et al., 2018; You et al., 2021). Recent studies have also explored the generalisability of machine learning models, such as RF, by training with the historical dataset from one area and subsequently applying the trained model to different regions across varying temporal scales and thereby assessing the model's robustness in accommodating variations in crop growth environments (Hao et al., 2020; Xu et al., 2020). However, traditional machine learning methods are not inherently designed to analyse the intrinsic spatial and temporal relationships present in multi-temporal satellite observations across crops' growing seasons. The extraction of meaningful temporal features from remote sensing time series, which represents the sequential relationships inherent in crop growth patterns, relies heavily on domain expertise and expertise for the development of handcrafted, predefined temporal features that accurately represent the characteristics of crop growth (Zhong et al., 2014; You and Dong, 2020).

Deep neural networks offer a significant improvement over these conventional classifiers in the task of crop mapping. Their architectural designs and feature extraction capabilities, especially in models like Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and their associated variants, make them particularly suited for processing sequential satellite data (Rußwurm and Körner, 2018; Xu et al., 2021b; Turkoglu et al., 2021; Zhong et al., 2019; Gallo et al., 2023; Liu et al., 2023). For example, the Long Short-Term Memory (LSTM) model (Hochreiter and Schmidhuber, 1997), a type of the typical RNN model, is developed for discovering time series data due to its ability to capture temporal dependencies and retain information over long sequences. Previous studies have utilised LSTM to process temporal observations obtained from multi-temporal satellite images across crop growth stages for crop mapping (e.g. Zhong et al., 2019; Xu et al., 2020; Dou et al., 2021). While LSTM-

based networks have demonstrated effectiveness in learning from time series remote sensing data, their performance is limited to the availability of substantial training datasets and computational costs (Xia et al., 2022). In contrast, RF could achieve higher classification accuracy than LSTM when using a limited number of training samples (Yuan and Lin, 2020). Another prominent network to handle long sequences is Transformer, which originated from natural language processing (Vaswani et al., 2017). It can process sequences in parallel and allows the model to focus on the most informative parts of the time series. Its application in crop mapping has been demonstrated by various studies (e.g. Rußwurm and Körner, 2020; Xu et al., 2021b).

Fully Convolutional Neural Networks (FCN) have also emerged as a powerful deep learning architecture, showcasing their ability to execute pixel-wise segmentation on multi-scale imagery in an end-to-end manner (Volpi and Tuia, 2017). Semantic segmentation, which entails assigning predefined labels to every pixel in an image, can leverage the FCN-based structures for crop mapping (Wei et al., 2019; Wei et al., 2021). Given the inherent 2D nature of remote sensing images, U-Net can adeptly extract high-level 2D spatial features by utilizing a cascade of convolutional filters through multiple nonlinear transformations (Ma et al., 2018). For instance, studies by Mohammadimanesh et al. (2019) and Wei et al. (2019) have employed the 2D U-Net for tasks like land cover classification and crop mapping, leveraging features such as backscattering, coherence, and polarimetric SAR (PolSAR). However, there's a growing consensus that the 3D U-Net, with its 3D convolution kernels, is capable of understanding the temporal dynamics of crop samples throughout their growth cycle and can extract spatiotemporal features from crop growth patterns over time in multi-temporal satellite imagery (Adrian et al., 2021; Gallo et al., 2023).

Although existing models, combining either optical or SAR-derived features, demonstrate their potential for extensive agricultural applications, most studies predominantly target regional-scale crop mapping within a single-year timeframe without considering the intricacies of smallholder-scale farmlands characterized by specific intercropping strategies, and interannual variations of cropland distribution within the same region. Additionally, model transferability, when applied to unseen data, remains challenging due to variations in climate, spectral signatures, topographical features, crop structures and agricultural practices (Lobell and Azzari, 2017; Hao et al., 2020; You et al., 2023). There is a need for a spatiotemporally generalizable classification scheme and a novel model architecture, tailor-made for interannual crop mapping

in regions exhibiting certain patterns like strip cropping. This study aims to design a transfer learning scheme of crop mapping in intercropping areas with smallholder croplands, using multiple models fed with multi-temporal Sentinel-1 coherence and multispectral bands from Sentinel-2 and RapidEye. Specifically, the study i) evaluates interannual and spatial generalisability of pre-trained models in the context of mapping strip cropping system, proposing a rule-based ensemble learning method that combines multi-source output probabilities using optimal thresholds; ii) investigates the synergistic use of multi-temporal InSAR coherence data and multispectral bands for enhanced predictive performance; and iii) interprets input feature importance based on learned features derived from the proposed model for multi-temporal crop mapping.

5.2 Materials

5.2.1 Study area

This study focused on Bei'an county, located in Heilongjiang province in northeast China (47°35'N ~ 48°33'N, 126°16'E ~ 127°53'E) (Figure 5-1), covering an area of 7,149 km^2. It is characterised by a humid continental monsoon climate, with an average annual temperature of 1.2°C and an average annual precipitation of 529 mm. These meteorological conditions, coupled with fertile soil, make this region suitable for the cultivation of spring maize, soybeans, wheat, rice, and other crops (Zhang et al., 2021). In Bei'an, the typical cropping routine involves sowing maize in late April, followed by soybean planted in early May, generally less than 10 days apart. The growing season for both crops is approximately four months in length, and crops are typically harvested in September. However, this schedule may vary annually due to local cropping practices, such as crop rotation strategies. In terms of land use, maize and soybean are the major crops in Bei'an, comprising 29.5% and 61.8% of the total sown area (2,190 km^2), respectively. Wheat is a less represented crop type and accounts for 2.9% of the area (Heihe Social and Economic Statistics Yearbook, 2018). To evaluate the performance of model transfer learning, two 10 x 10 km sites (labelled as A and B) were selected in 2018, both exhibiting complex intercropping patterns.

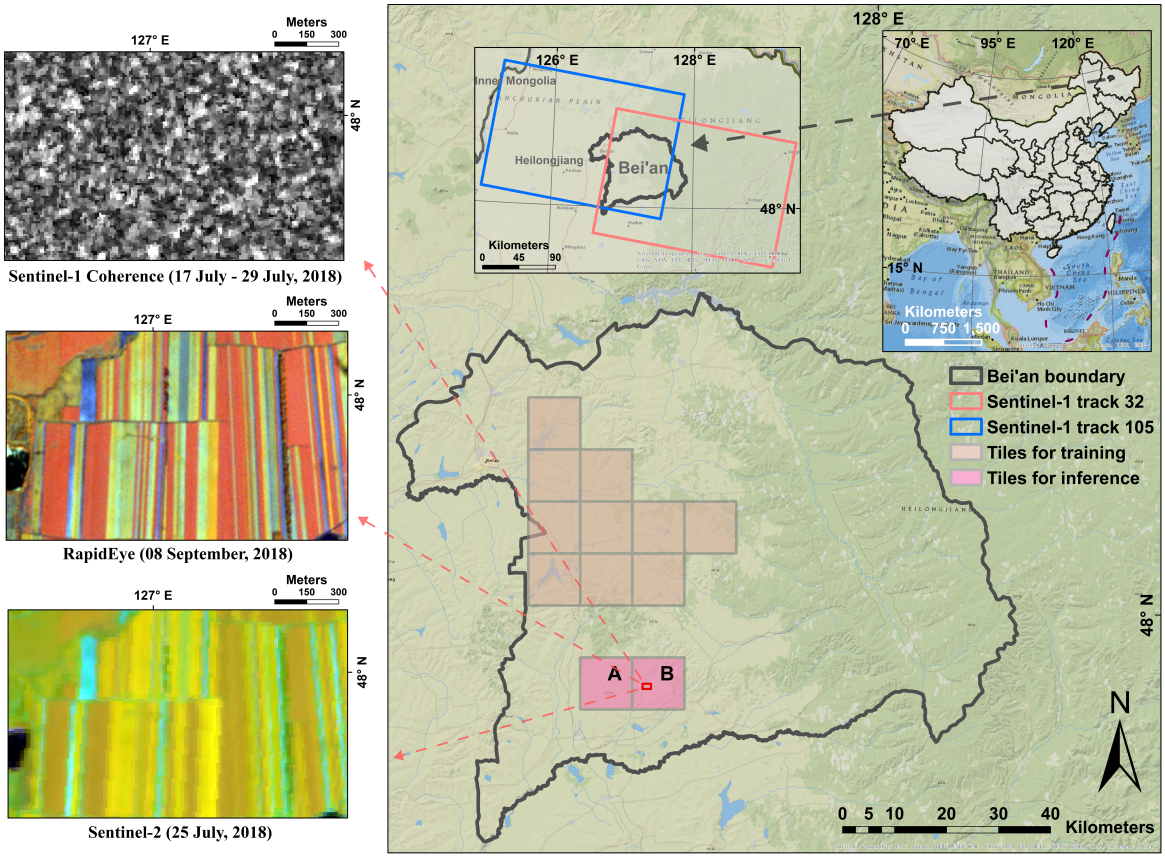

Figure 5-1. The location of Bei'an and spatial distribution of the designated training/inference tiles within the study area. Each tile is a 10 km by 10 km grid at 5 m resolution (2000 × 2000 pixels). The coordination system for inset maps of Bei'an is EPSG:32652 - WGS 84 / UTM zone 52N. Inset maps: Sentinel-1 coherence (VV) generated between the 17th and 29th July 2018 acquisitions. RapidEye with R: Near-Infrared (NIR), G: Red Edge, B: Red. Sentinel-2 with R: B8a (Vegetation Red Edge), G: B11 (SWIR), B: B4 (Red).

5.2.2 Satellite datasets and pre-processing

In this study, the image collection contained 23 Sentinel-1 scenes from 2017 and 22 from 2018. This was further supplemented by three Sentinel-2 acquisitions and a single RapidEye acquisition for each year (Figure 5-2). The timing of the multi-sensor data collection was in accord with the local sowing routines and the crops' growth cycles, which ranged from early May to late September.

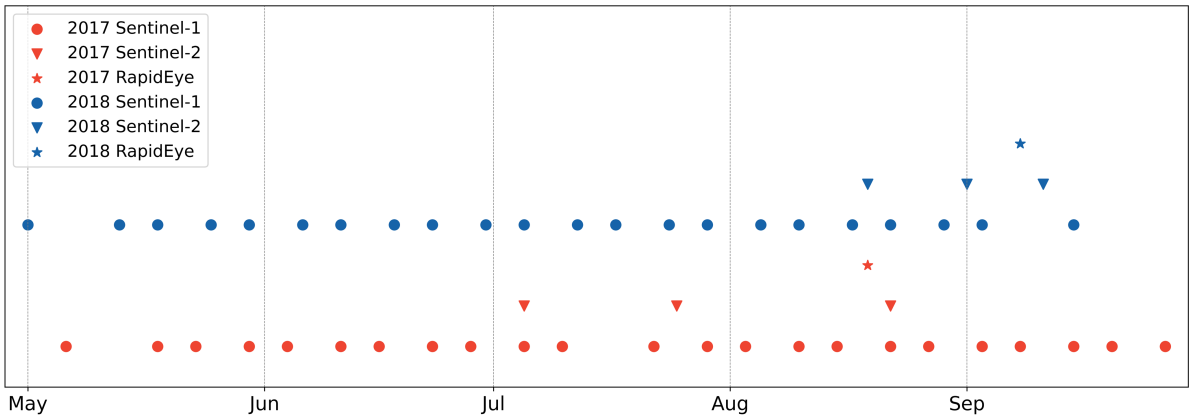

Figure 5-2. Multi-source satellite acquisition collection covering the study area in 2017 and 2018.

5.2.2.1 Sentinel-2 and RapidEye datasets

The multispectral bands used in this study were obtained from Sentinel-2A/B Level-1C (L1C) products and the RapidEye Ortho (Level-3A) product (See Table 5-1). The Sentinel-2 L1C data were transformed to Level-2A for atmospheric correction and retrieving bottom-of-atmosphere (BOA) reflectance values. This was completed using the 'Sen2Cor' tool (Main-Knorn et al., 2017) within the Sentinel Application Platform (SNAP 8.0) developed by the European Space Agency (ESA). Sentinel-2 datasets from 2017 to 2018 were selected based on the average cloud cover not exceeding 8%. RapidEye imagery, orthorectified as individual tiles of 25 × 25 kilometres with a spatial resolution of 5 meters, were atmospherically and topographically corrected using ERDAS IMAGINE 16.5 to obtain surface reflectance values. The selection of optical bands was informed by previous studies (Cai et al., 2018; You and Dong, 2020) that demonstrated their sensitivities in differentiating maize and soybean, particularly in Northeast China. All optical bands utilized in this study were resampled to 5 m spatial resolution.

Table 5-1. Summary of the optical bands used in this study.

Sensor	Bands	Central Wavelength (nm)	Resolution (m)
Sentinel-2	B4 – Red	665	10
	B8a – Vegetation Red Edge	865	20
	B11 – SWIR	1610	20
RapidEye	Red	630 – 685	5
	Red Edge	690 – 730	5
	NIR	760 – 850	5

5.2.2.2 Sentinel-1 coherence

Multitemporal Sentinel-1B Single Look Complex (SLC) C-band data, obtained in Interferometric Wide (IW) swath mode, were used to calculate coherence. These datasets have

a spatial resolution of 10 meters and a 12-day repeat cycle. Pre-processing steps used to derive coherence followed those in Nasirzadehdizaji et al. (2021) and Donezar et al. (2019). The InSAR processing was performed using the Sentinel-1 Toolbox in SNAP 8.0. Coherence maps were calculated based on image pairs from the same tracks (for both track 105 and 32), maintaining a minimum 12-day temporal baseline in both VV (vertical transmit and vertical receive) and VH (vertical transmit and horizontal receive) polarizations to construct the coherence matrices for 2017 and 2018. Each track generated 10 coherence maps to ensure consistency in baselines between consecutive images. It is important to adhere to the same orbit path for coherence calculation. The coherence values in the resulting maps range from 0 to 1, with 1 indicating complete coherence between images, and 0 representing no coherence. These coherence values are influenced by temporal changes in the scattering characteristics of the observed targets, as noted by Ferretti et al. (2007). For analysis with the multispectral datasets, coherence maps were resampled to 5 m spatial resolution.

5.2.3 Reference data

In this study, reference data were used to train and evaluate the performance of the models. The distribution of cropland across two consecutive years, 2017 and 2018, is illustrated in Figure 5-3. The field polygons were surveyed in 2017 by the Chinese Academy of Agricultural Sciences (CAAS). Narrow strip-like field parcels were classified using higher-resolution imagery, such as RapidEye, whereas larger crop parcels were manually labelled based on Sentinel-2 imagery. Similar procedures were conducted for 2018. The variations observed in the cropland distributions between these two years can be attributed to interannual crop rotation practices. For training, the study incorporated a Cropland Data Layer (CDL) predicted by Liu et al. (2023) for Bei'an in 2017. Since the CDL lacked a confidence threshold for trusted pixels, the predicted pixels within the CDL were replaced by the reference data from 2017, following the alignment in the geolocations of the corresponding pixels. The remaining predicted pixels were retained in the CDL. Notably, no CDL mask was used for 2018.

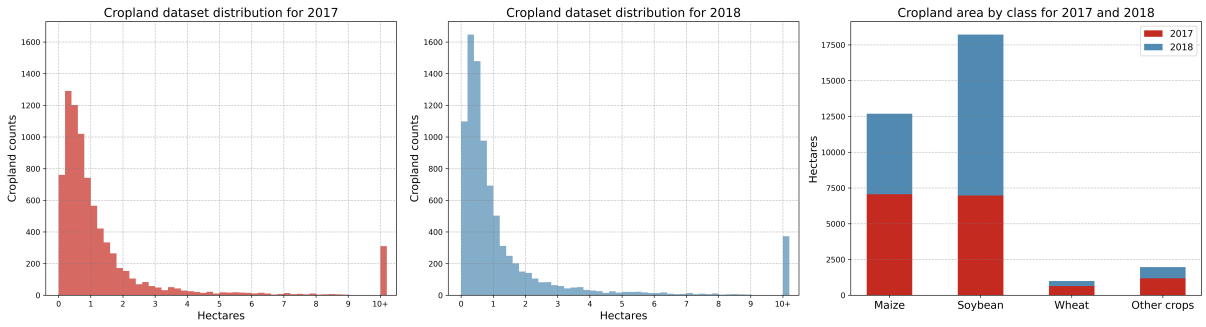

Figure 5-3. Distribution frequency of cropland sample sizes for 2017 (left) and 2018 (centre), followed by the comparison of the cropland sample areas segregated by crop categories for both years (right).

5.3 Methods

The general workflow for this study is presented in Figure 5-4. The data from the three sensors were pre-processed for generating the input dataset. Ground-truth samples were collected from 2017 and 2018, along with the Cropland Data Layer (CDL) for 2017, to create the training set. The training tiles in Figure 5-1 were divided into subsets of tiles, with 60% used for training, 20% for validation and 20% for testing. The classifiers were trained using the training set and then applied to Sites A and B in 2018 to map the cropland area. The transfer learning included direct prediction using pre-trained models. It also included indirect prediction based on fine-tuning with samples from 2018 to enhance the model's generalisability and adaptability to spatiotemporal variations within a region. Finally, the feature importance analysis was conducted to assess the contribution of time series multi-source inputs to the classification accuracy.

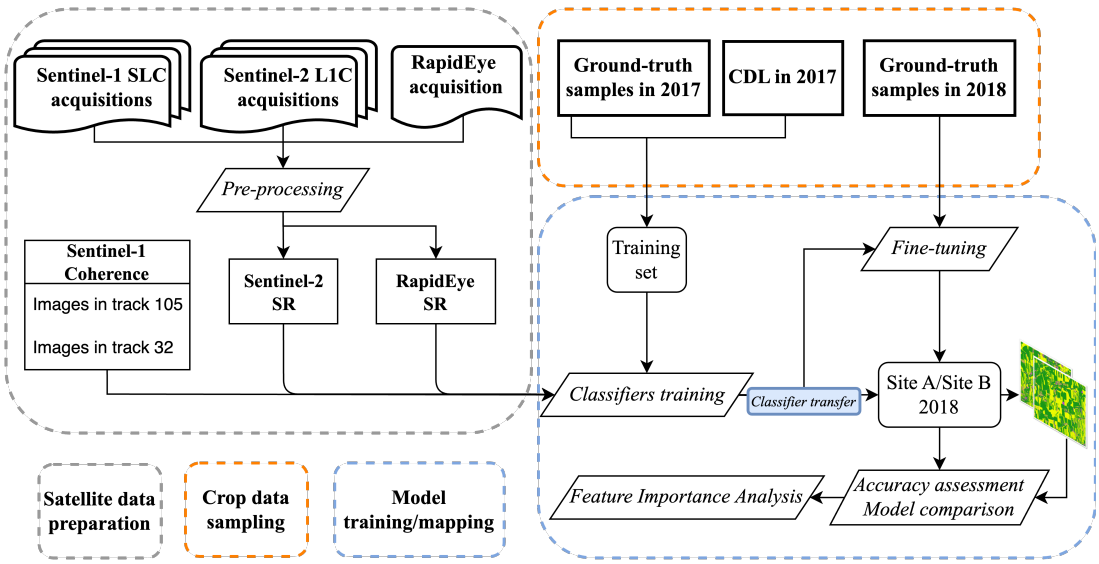

Figure 5-4. The general workflow of this study. SR stands for surface reflectance.

5.3.1 InSAR coherence estimation

InSAR coherence measures the similarity of the interferometric phase between two co-registered SAR acquisitions obtained at different times (Ferretti et al., 2007), which indicates the normalized cross-correlation coefficient between complex Sentinel-1 SLC image pairs used in this study. The expression of its absolute value (Touzi et al., 1999) can be described using Eq. (5-1):

$$\gamma = \frac{\left|\frac{1}{N}\sum_{i=1}^{N} S_{1i}S_{2i}^{*}\right|}{\sqrt{\left(\frac{1}{N}\sum_{i=1}^{N} S_{1i}S_{1i}^{*}\right)\left(\frac{1}{N}\sum_{i=1}^{N} S_{2i}S_{2i}^{*}\right)}} \tag{5-1}$$

where γ is coherence between two complex co-registered SLC acquisitions that contain the master image S_1 for reference and slave image S_2 for repeat, and * denotes that one of the images is conjugated. The magnitude of the complex sum is normalized by dividing by a normalization factor (the denominator), ensuring that the magnitude of γ falls within the real number range between 0 and 1. Images with low coherence values indicate decorrelation due to spatial and temporal decorrelations, and system noise (Nasirzadehdizaji et al., 2021). This is particularly important during the crop growth seasons, where rapid changes in surface scatterers highlight the impact of temporal decorrelation. Conversely, high coherence values suggest uniformity in the physical properties or scatterings' position in image pairs over time. We considered a subset of the interferometric combinations from all 20 image pairs annually, each pair with the shortest 12-day temporal baseline (See Figure 5-5). The coherence matrix's main diagonal entries represent the shortest temporal baseline for two consecutive images. i.e., multi-track Sentinel-1 SAR image pairs in VV and VH polarization were derived based on 12-day intervals. Furthermore, the coherence estimation was segregated by different orbits (tracks) due to the variations in viewing angles and capture conditions resulting from the different sensor paths. These factors can lead to coherence measurement disparities between tracks that cover overlapping ground areas within the study region.

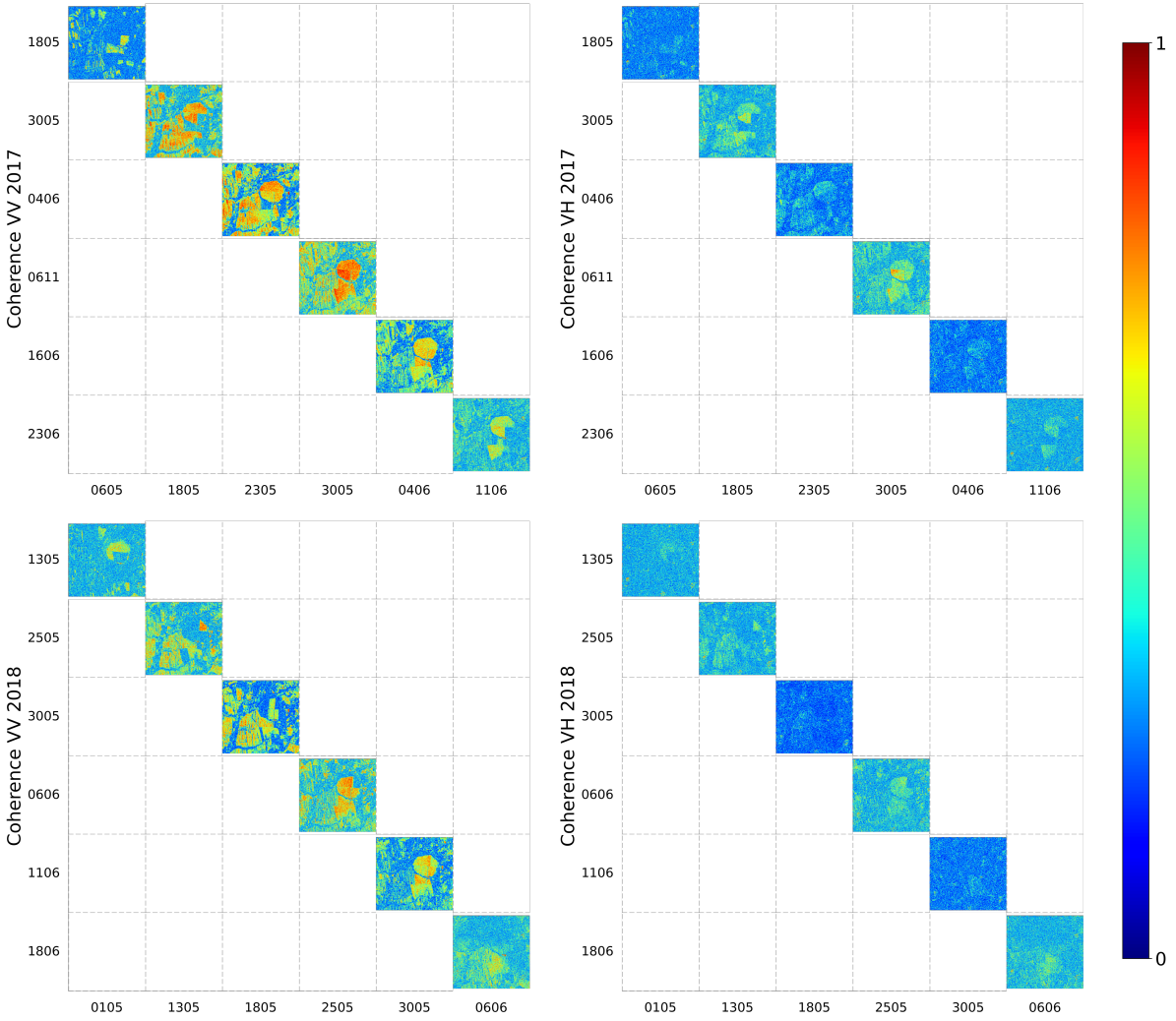

Figure 5-5. InSAR coherence matrix maps for the VV and VH bands from multi-track image pairs for 2017 and 2018. These maps are plotted along the X and Y axis (displayed in the day-month format), representing each image pair from each track. Note that only the first 6 image pairs are shown for better visualisation. The sequence of plots, from left to right, corresponds to tracks in the order of 32, 32, 105, 32, 105, 32, for each year of polarisations.

5.3.2 Classification models

Different deep learning model architectures were explored (See Figure 5-6). These are described in further detail below.

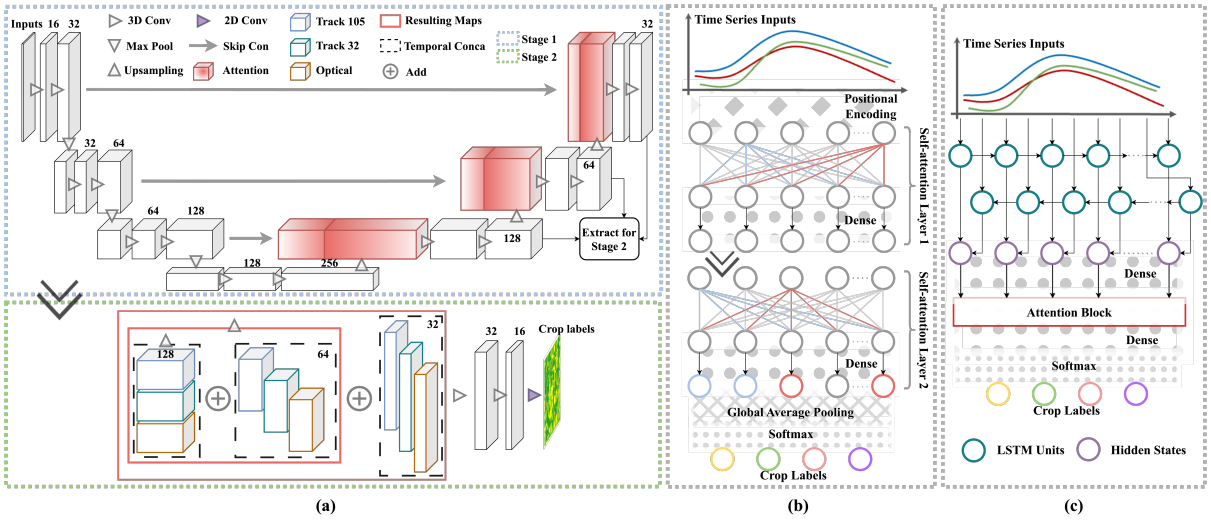

Figure 5-6. The model architectures of 3D U-Net (a), Transformer (b) and AtLSTM (c). In (a), '3D Conv' and '2D Conv' represent the three-dimensional/two-dimensional convolutional process. 'Skip Con' refer to the skip connection. 'Temporal Conca' is the time series concatenation. In (b) and (c), each processes multi-source inputs in parallel; the outputs from each source are concatenated channel-wise, followed by the Softmax function for predicting the classes.

5.3.2.1 Attention 3D U-Net

An attention-based 3D U-Net framework (Figure 5-6 (a)) is designed to effectively process volumetric data. This novel architecture integrates an attention mechanism that selectively emphasizes salient features and suppresses less relevant regions in the feature map. This improves the model's ability to focus on important spatiotemporal input features. The basic 3D U-Net architecture follows Çiçek et al. (2016), utilizing 3D convolutions with batch normalization and ReLU activations to maintain the representational integrity of the spatial information. The model employs a symmetric design with downsampling and upsampling pathways, ensuring detailed feature extraction and dimensional recovery. The downsampling path consists of consecutive convolutional blocks followed by max pooling to reduce spatial dimensions, whereas the upsampling path utilizes transpose convolutions combined with attention blocks to effectively recover spatial details. Within this architecture, the attention block serves as a gating mechanism to refine the feature fusion process (Oktay et al., 2018). This gating mechanism enables the network to selectively focus on the most relevant features from the input data, which can be generally formulated as Eq. (5-2):

$$\hat{x} = \sigma(ReLU(W_x * x + W_g * g + b_x + b_g) * W_f + b_f) * x \quad (5\text{-}2)$$

where the input matrix x and the gating signal g are linearly transformed through 3D convolution operation with respective weight matrix W_x and W_g, followed by the nonlinear

activation $ReLU$. Then the attention coefficient is derived by the sigmoid activation σ of the 3D convolution with W_f performed on their combined feature map. The attention output is then calculated by the element-wise multiplication, denoted by *, between x and the coefficient. b_x, b_g and b_f represent the bias vectors corresponding to each convolution process. Each 3D convolution has kernels with size 1×1×1. The gating signal g indicates the intermediate feature map produced during the downsampling stage; it is used to refine the feature map coming from the lower-resolution layer during the upsampling process.

This multi-source architecture facilitates the integration of data sources by processing them through parallel pathways, each pathway being equipped with an attention 3D U-Net architecture. At each level of the upsampling blocks, the feature maps generated from each data source are initially concatenated separately and then combined level by level via element-wise addition (Tao et al., 2022), as shown in Stage 2 of Figure 5-6 (a). This approach allows the model to learn more robust feature representations by leveraging information from multiple sources. The final output of the model is derived through a sequence of convolutions, pooling, and softmax functions implemented in the subsequent layers, which progressively reduces the dimensionality of the data to yield the final prediction map. This structure ensures that the model effectively synthesizes the input data, using the strengths of each data modality to enhance the overall predictive performance.

5.3.2.2 Transformer

The Transformer network utilised self-attention mechanisms to calculate the correlations between every pair of embedding vectors across all the time steps (Figure 5-6 (b)), which enables the effective extraction of temporal dependencies, even within very long sequences. The self-attention is operated by using three vectors (*Query*, *Key* and *Value*) that originate from three linear transformations applied to the same input vectors. The output of self-attention for each position (SA_i) can be expressed as Eq (5-3):

$$SA_i = \sum_{j=1}^{N} Softmax\left(\frac{(W_Q x_i)\cdot(W_K x_j)^{\intercal}}{\sqrt{D_K}}\right)\cdot W_V x_j \tag{5-3}$$

where W_Q, W_K and W_V are the weight vectors used for mapping input vectors x_i and x_j to corresponding *Query*, *Key* and *Value* vectors. The dot product of *Query* at position i *and Key* at position j is the attention score $(W_Q x_i)\cdot(W_K x_j)^{\intercal}$ between two positions in a sequence N,

which is normalised by the square root of the hidden dimension of the *Key* vectors $\sqrt{D_K}$. Then, the *Value* vectors ($W_V x_j$) are multiplied by the Softmax-scaled attention scores to derive a weighted sum of the *Value* vectors for each position, which represents the self-attention output. This output is an enhanced representation of each input that incorporates information from the most relevant parts of the input sequence as identified by the self-attention mechanism.

In this study, the standard Transformer architecture, initially introduced by Vaswani et al. (2017), has been adapted, following the approach of Rußwurm and Körner (2020). This modification involves the incorporation of a positional encoding function into the time series inputs. This addition enables the self-attention mechanism to effectively process the sequential correlation inherent in time series data. The positionally encoded time series data are then processed through stacked multi-head self-attention layers. The multi-head self-attention process is replicated across multiple 'heads', each operating in parallel and equipped with its own set of vectors W_Q, W_K and W_V. This arrangement allows the model to capture various dependencies within a data sequence. The outputs from all heads are then concatenated and linearly transformed to produce the final output. This is followed by a global average pooling across the temporal dimension to achieve a more condensed data representation. Similar to the attention 3D U-Net, this study's Transformer-based architecture is designed to accommodate multi-source inputs. The outputs corresponding to each input source are concatenated, followed by fully-connected layers with the Softmax function for predicting crop types.

5.3.2.3 AtLSTM

Another temporal architecture adding an attention mechanism after bidirectional Long Short-Term Memory units (AtLSTM), was employed in this study (Figure 5-6 (c)). This design enables the recurrent neural networks to specifically focus on certain segments of the input sequence during the output prediction process. It utilises bidirectional LSTM layers to process the input sequence in both forward and backward directions, which is particularly effective in capturing data's temporal dynamics, as it concatenates hidden states derived from both directions. The attention weights are calculated using the hidden states from bidirectional LSTM layers as follows in Eq. (5-4):

$$A_t = Softmax((W_a H_t + b_a) \cdot h_T) \tag{5-4}$$

where W_a is the weight matrix of a dense layer that transforms the sequence of hidden states H_t from bidirectional LSTM into a new feature dimension $W_a H_t$ with bias b_a. Then the attention score is given by the dot product between the computed component and the last hidden state h_T, which then were normalised by the Softmax function to produce attention weights for each time step A_t. In the final stage of the AtLSTM process, the attention outputs are derived from a weighted sum of the hidden states, calculated using the attention weights. These outputs are then concatenated with the last hidden state, which serves to enhance the model's decision-making capabilities.

In this study, the AtLSTM was utilized as an auxiliary model in conjunction with the Transformer network. The critical assessment of model transfer learning performance was not conducted independently for AtLSTM. Instead, its contribution was evaluated in the context of its integration with the Transformer model, focusing on the synergistic effect of combining these architectures.

5.3.2.4 Decision fusion of Transformer-AtLSTM-RF

The Transformer-AtLSTM-RF framework proposed in this study leverages the power of ensemble learning by combining the capabilities of the Transformer and AtLSTM networks. The probability outputs from these models are integrated with those derived from the Random Forest (RF) classifier. The final probabilities are determined using a rule-based decision fusion method. The ensemble approach (Transformer-AtLSTM) is designed as a two-branch architecture, where each branch processes multi-source inputs in parallel. Specifically, each classifier within the ensemble is trained independently, and their outputs are then combined and normalized through a Softmax function to derive probabilities.

Another component of this framework is the RF model, which utilizes decision-tree classifiers based on the bagging strategy. This approach enhances model generalizability by efficiently managing a large number of input variables and identifying the best split with relatively low computational complexity (Breiman, 2001). RF has been widely established as a baseline model in crop mapping studies (Zhong et al., 2019; Rußwurm and Körner, 2020; Xu et al., 2021b; Turkoglu et al., 2021) and is known for its ability to estimate probabilities associated with predictions by counting the proportion of trees that vote for each class.

In this ensemble framework, the outputs from Transformer-AtLSTM (Ensemble) and RF are two vectors of probabilities. Each vector consists of channels corresponding to the number of classes, with each channel element representing the likelihood of the input belonging to a specific crop class. These probability outputs are then combined based on a predefined threshold applied in the rule-based decision fusion method (Li et al., 2019), as follows:

$$P_{i,j}^{*} = \begin{cases} P_{i,j}^{1} & if\ P_{i,j}^{1} \geq \alpha \\ P_{i,j}^{2} & otherwise \end{cases} \tag{5-5}$$

where $P_{i,j}^{1}$ is the output probability from Transformer-AtLSTM (Ensemble) on sample i and class j, and $P_{i,j}^{2}$ stands for probabilities from RF. α is the optimal threshold determined using a grid search (Hsu et al., 2003). $P_{i,j}^{*}$ indicate the probabilities from both models, which are filtered through the threshold. This effective ensemble synergy merges the outputs of individual classifiers based on a specific fusion rule, which capitalises on the complementary behaviours of different models to enhance classification performance (Clinton et al., 2015).

5.3.3 Model Implementation

The structure of input data in this study varied according to the model architectures employed. For the 3D U-Net model, image tiles from different sources used for training were randomly cropped into 128 × 128 patches. To ensure sufficient semantic context, thresholds for the minimum proportion of non-zero (non-background) labels within a patch were set at 50% for 2017 and 70% for 2018. The increase in the threshold from 2017 to 2018 implies that patches in 2018 contained a larger amount of unlabelled pixels compared to those in 2017. These adjusted thresholds reflect the requisite proportion of significant label data necessary in an image patch to be used for model training. Additionally, data augmentation techniques such as random scaling, vertical and horizontal flipping, and rotation were applied on the fly to image patches during the training process. The input shape for 3D U-Net includes batch, height, width, time, and channels. The image patches were separated by coherence from two tracks and the optical source. Regarding pixel-based inputs for the Transformer and AtLSTM, data with no labels and missing data were excluded. The data structure for these models consists of batch, time, and channels for each data source. The RF, while sharing the same dataset as the Transformer and AtLSTM, stacked data from all sources and reshaped the structure into a one-

dimensional vector including batch and channels. This adaptation was necessary because RF, unlike the other temporal models, is not designed to process the temporal dimension of data.

The optimal model hyperparameters were determined through search spaces of candidate values. For example, the dimension of the self-attention layer was set at 128 in {64, 128, 256}. The number of heads was 4 in {4, 6, 8}. The number of layers was 2 in {1, 2, 4}. The last dens layer had 512 units in {128, 256, 512, 1024}. For AtLSTM, the number of LSTM units was set to 128 in {64, 128, 256}, and one layer was used from {1, 2}. The dimension of the hidden features in the attention block followed the dimension size of output from the bidirectional LSTM, which doubled the number of unilateral LSTM units. The number of trees for RF was set to 500 from {200, 400, 500}, and the best split was set to 4 from {2, 4, $\sqrt{Time\ steps\ \times\ bands}$}. For Transformer-AtLSTM-RF, the optimal threshold α in the grid search was found to be 0.51 in Site A, and 0.64 in Site B. Within the Transformer-AtLSTM (Ensemble), only the Transformer branch was pre-trained with data in 2017. The AtLSTM functioned as an auxiliary model during the fine-tuning phase. Additionally, the study explored the use of the pre-trained Transformer as a backbone module, linked with AtLSTM in a hybrid configuration termed Transformer-AtLSTM (Backbone).

Regarding the training configuration, weighted cross-entropy loss and the Adam optimizer with a learning rate of 0.001 were applied for deep learning models. A lower learning rate of 0.00001 was utilized during fine-tuning with new data to preserve the knowledge of prior training. We experimentally assessed the number of 'frozen' layers in pre-trained models, which are characterized by learnable weights that are fixed during the fine-tuning process. By fine-tuning the 'unfrozen' layers with a lower learning rate, the models could better adapt to the new dataset while retaining prior knowledge from pre-training. Furthermore, all deep learning models in this study were implemented in TensorFlow (2.14.0) on Google Colab under GPU A100. The RF was developed using the Scikit-learn package (1.3.0) in Python (3.7.15) on two Intel (R) Xeon (R) Silver 4114 CPU processors (2.20GHz/2.19 GHz). For the evaluation of classification performance, metrics such as overall accuracy (OA), mean F1 score (F1), and mean intersection over union (mIoU) were utilized in the experiments.

5.4 Results

5.4.1 Coherence temporal profile

The temporal profiles of crops' InSAR coherence levels, based on polarisations and tracks for the years 2017 and 2018, are illustrated in Figure 5-7 and Figure 5-8. Generally, for both crops, the coherence value peaks around the end of May and early June, post-seeding. The value starts to decrease as the crops grow. In 2017, maize showed a noticeable drop from the 'jointing to tasseling' to 'silking' stage. Similarly, high coherence is observed for soybean during the emergence, which drops significantly during 'first to sixth trifoliolate leaves'. It then stabilises and gradually declines through 'podding' to 'harvesting'. Wheat displayed the lowest coherence value during 'flowering', but a sharp increase was observed during 'harvesting'. The coherence trends for other crops were generally similar to wheat, with consistently low values from late June to early September. The year 2018 showed similar coherence patterns across all crop types, though overall coherence levels were lower compared to 2017, particularly evident during the wheat 'harvesting' stage.

Between the VV and VH polarisations from each track, VV consistently presented higher coherence values than VH. This difference is likely due to the sensitivity of co-polarization to enhanced volume scattering of vegetation and the reduced canopy penetration of VV polarization (Manavalan, 2018). Furthermore, VV and VH polarisations in track 32 maintained higher coherence throughout the crops' stages compared to track 105, which could result from sensor viewing geometries and spatial differences in the observed areas. This observed disparity in coherence between different polarisations and tracks could highlight the potential importance of capturing the structural information of crops throughout the season, which contributes to crop identification.

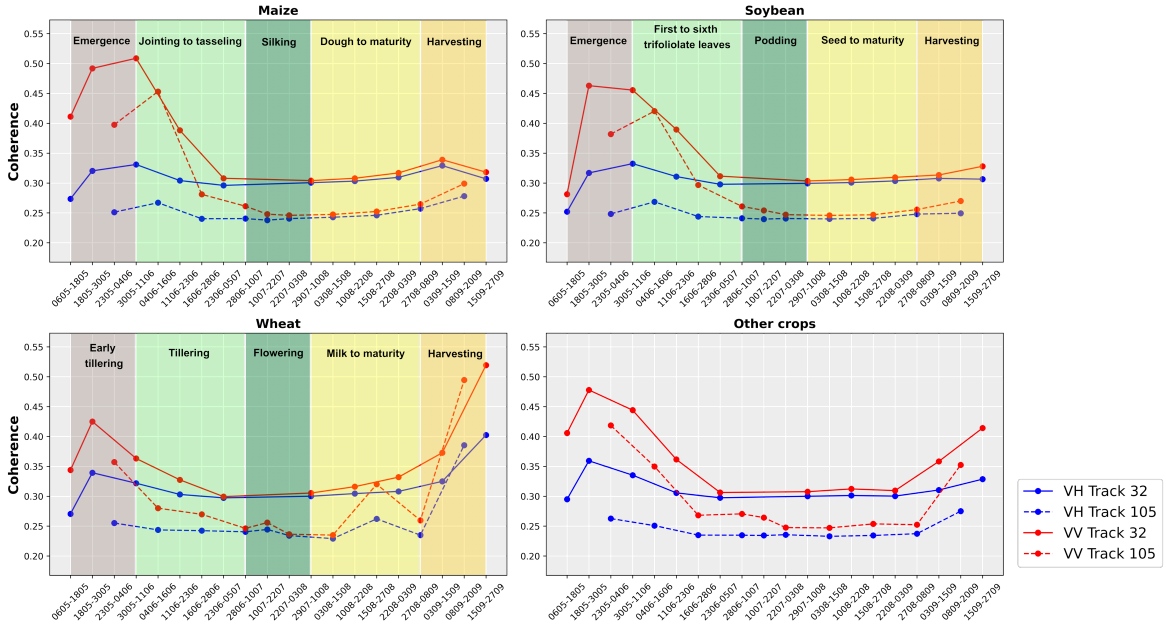

Figure 5-7. The 2017 temporal profile of mean coherence in VH and VV bands separated by tracks.

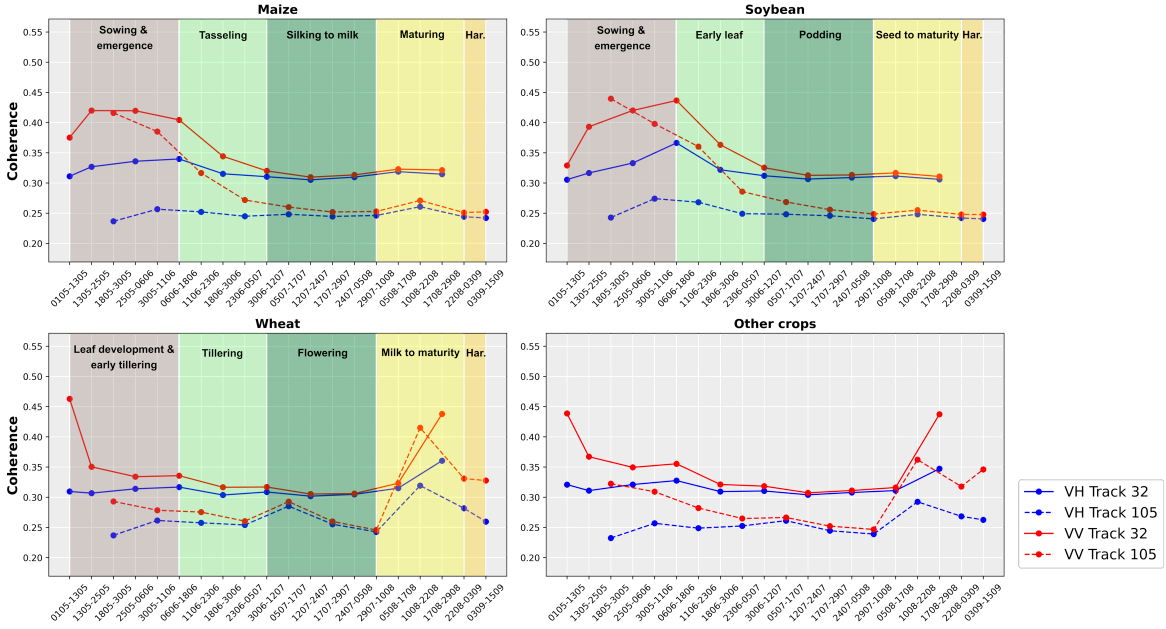

Figure 5-8. The 2018 temporal profile of mean coherence in VH and VV bands separated by tracks. 'Har.' refers to the harvesting stage.

5.4.2 Transfer learning accuracies for the sites A and B in 2018

In Table 5-2 and Table 5-3, the Transformer-AtLSTM-RF model achieved the highest mean F1 score at 92.7%, alongside the highest OA of 96.2% and mIoU of 86.9%. Analysing performance by individual crop categories, the Transformer-AtLSTM-RF model outperformed other models for maize (F1:96.0%, IoU: 92.3%), soybean (F1:96.7%, IoU: 93.6%) and other

crops (F1: 85.5%, IoU: 74.7%), despite the marginal drop for maize compared to the Transformer-AtLSTM (ensemble). The RF, trained specifically with 2018 data, yielded a higher F1 score (81.7%) and IoU (69.1%) for other crop categories than the Transformer-AtLSTM (ensemble), which achieved an F1 score of 77.4% and an IoU of 63.1%. This contributed to the improved identification of other crops in the Transformer-AtLSTM-RF ensemble (F1: 85.5%, IoU: 74.7%). It is notable that the pre-trained 3D U-Net and Transformer models, without fine-tuning, demonstrated relatively lower classification performance across all metrics, particularly for other crops. This indicates the significant improvement in classification accuracy that can be achieved through the fine-tuning process.

Table 5-2. Transfer Site A: overall accuracy (OA) and mean F1 score for 2018 crop mapping validation. '-': no fine-tuning applied, '*': fine-tuned with 2018 data. Columns with bold values indicate the best performance.

Model		Fine-tuning	Maize (%)	Soybean (%)	Other crops (%)	Mean F1 (%)	OA (%)
3D U-Net		-	73.7	68.3	1.3	47.8	61.7
		*	87.8	91.1	69.2	82.7	84.3
Transformer		-	75.4	54.0	3.2	44.2	50.2
		*	95.0	95.7	54.9	81.9	94.5
Transformer-AtLSTM	Backbone	*	95.2	95.7	56.7	82.5	94.3
	Ensemble	*	**96.1**	**96.7**	77.4	90.1	96.1
RF pre-trained		-	80.5	81.1	1.5	54.3	68.5
RF trained with 2018 data		-	94.8	95.8	81.7	90.8	95.1
Transformer-AtLSTM-RF		-	96.0	**96.7**	**85.5**	**92.7**	**96.2**

Table 5-3. Transfer Site A: IoU and mean IoU (mIoU) of 2018 crops. '-': no fine-tuning applied, '*': fine-tuned with 2018 data. Columns with bold values indicate the best performance.

Model		Fine-tuning	Maize (%)	Soybean (%)	Other crops (%)	mIoU (%)
3D U-Net		-	58.4	51.8	0.7	37.0
		*	78.3	83.6	53.0	71.6
Transformer		-	60.5	37.0	1.6	33.0
		*	90.5	91.8	37.8	73.4
Transformer-AtLSTM	Backbone	*	90.8	91.7	39.5	74.0
	Ensemble	*	**92.6**	**93.6**	63.1	83.1
RF pre-trained		-	67.4	68.2	0.7	45.4
RF trained with 2018 data		-	90.0	91.9	69.1	83.7
Transformer-AtLSTM-RF		-	92.3	**93.6**	**74.7**	**86.9**

The transfer learning performance for Site B is presented in Table 5-4 and Table 5-5. Similar to the results from Site A, the Transformer-AtLSTM-RF model achieved the highest mean F1 (88.6%), OA (90.7%) and mIoU (79.7%). This model's performance was notably enhanced by the decision-fusion technique, which combined the RF model trained from scratch with 2018

data and the Transformer-AtLSTM (ensemble), leading to the highest F1 and IoU for minor crops at this site. E.g., wheat (F1: 84.4%, IoU: 73.0%) and other crops (F1: 88.8%, other crops: 79.8%). However, it's important to note that this method resulted in slightly reduced performance for major crops like maize, where the F1 score and IoU were lower by 0.3% and 0.6%, respectively, compared to the highest accuracy values. Additionally, the fine-tuning of models with 2018 data demonstrated improved outcomes for 3D U-Net and Transformer at Site B, when compared to the transfer learning results without fine-tuning.

Table 5-4. Transfer Site B: overall accuracy (OA) and mean F1 score for 2018 crop mapping validation. '-': none fine-tuning applied, '*': fine-tuned with 2018 data. Columns with bolded values indicate the best performance.

Model		Fine-tuning	Maize (%)	Soybean (%)	Wheat (%)	Other crops (%)	Mean F1 (%)	OA (%)
3D U-Net		-	61.2	54.5	21.2	2.1	34.8	42.7
		*	71.7	81.5	76.0	85.8	78.7	70.4
Transformer		-	75.4	57.4	33.2	3.2	42.3	49.4
		*	85.8	89.8	58.5	66.0	75.0	84.9
Transformer-AtLSTM	Backbone	*	85.7	90.3	62.5	68.7	76.8	85.5
	Ensemble	*	**88.5**	92.6	78.7	84.9	86.2	89.9
RF pretrained		-	62.2	63.9	12.8	36.5	43.9	48.9
RF trained with 2018 data		-	86.0	92.1	73.6	71.4	80.8	87.9
Transformer-AtLSTM-RF		-	88.2	**93.0**	**84.4**	**88.8**	**88.6**	**90.7**

Table 5-5. Transfer Site B: IoU and mean IoU (mIoU) of 2018 crops. '-': none fine-tuning applied, '*': fine-tuned with 2018 data. Columns with bolded values indicate the best performance.

Model		Fine-tuning	Maize (%)	Soybean (%)	Wheat (%)	Other crops (%)	mIoU (%)
3D U-Net		-	44.1	37.5	11.9	1.0	23.6
		*	55.9	68.7	61.3	75.1	65.3
Transformer		-	60.5	40.2	19.9	1.7	30.6
		*	75.1	81.5	41.4	49.3	61.8
Transformer-AtLSTM	Backbone	*	74.9	82.4	45.5	52.3	63.8
	Ensemble	*	**79.4**	86.2	64.8	73.7	76.0
RF pre-trained		-	45.2	47.0	6.9	22.3	30.3
RF trained with 2018 data		-	75.5	85.4	58.3	55.5	68.6
Transformer-AtLSTM-RF		-	78.8	**86.9**	**73.0**	**79.8**	**79.7**

The visual comparisons of transfer learning performance for Site A and Site B are shown in Figure 5-9 and Figure 5-10, respectively. The 3D U-Net model exhibited a notable number of misclassifications at both sites, as indicated by the red patches, aligning with the results presented in Tables 5-2 – 5-5. The Transformer and RF successfully identified most crop parcels, but they experienced occasional misclassifications in local regions and areas with mixed crops, especially near the edges. The Transformer-AtLSTM models, in both combined configurations, demonstrated improvements in classification performance, refining areas that

were previously misclassified by the Transformer alone. Notably, the Transformer-AtLSTM-RF exhibited the best performance at both sites, aligning most closely with the ground-truth labels compared to other evaluated models. For more detailed visualization, Figure S7 and Figure S8 in the Supplementary material provide the enlarged scale of the intercropping areas in Sites A and B. Additionally, the confusion matrices for both sites, from the Transformer-AtLSTM-RF, are displayed in Figure 5-11. Both Site A and Site B show high positive prediction rates for major and minor crops. Site A's predictions are more accurate than Site B's, particularly for maize and soybean, probably due to the larger sample size for these crop types at Site A. Notably, wheat, as a minor crop type, showed the highest prediction accuracy of 96.7% (28,541) at Site B, demonstrating the model's robustness in diverse crop scenarios.

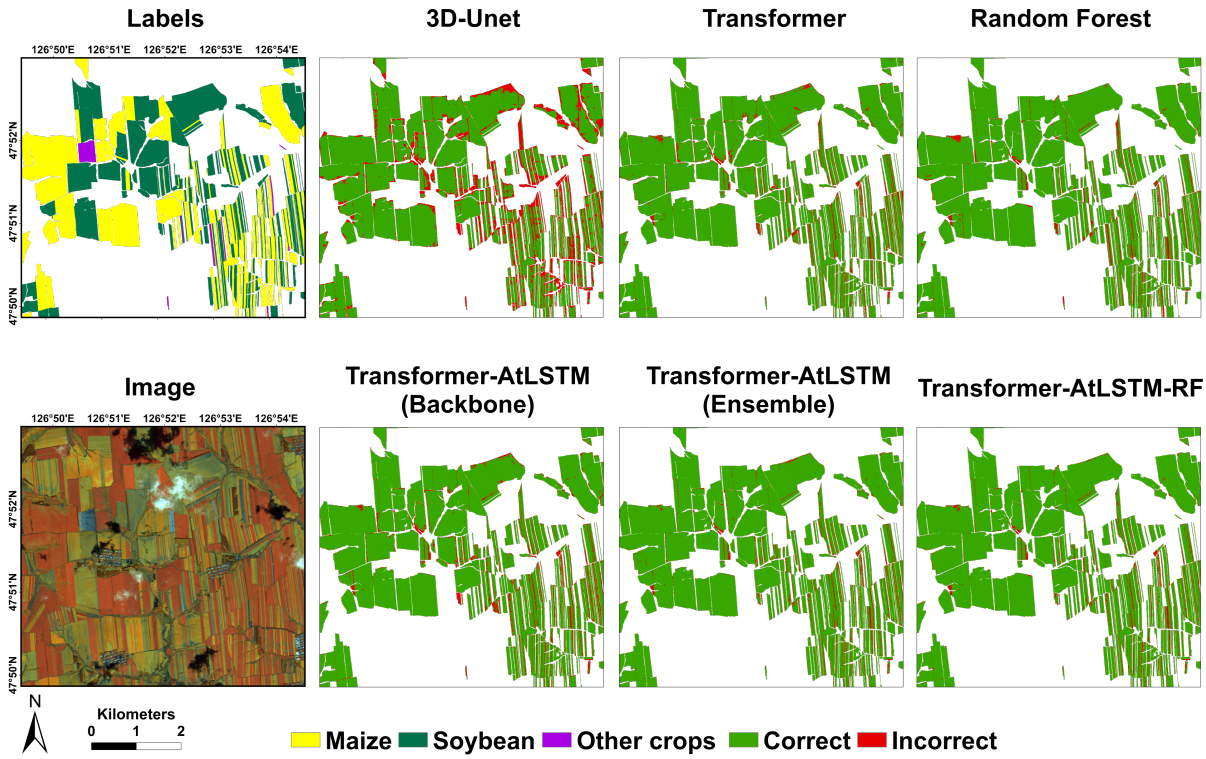

Figure 5-9. Crop mapping results for Site A in 2018. The difference maps are compared with ground-truth labels. Correctly classified pixels are shown in green, while misclassified pixels are highlighted in red. All deep learning-based models were fine-tuned with 2018 data. Random Forest (RF) was trained from scratch using 2018 data. The satellite image for Site A is a RapidEye false colour composite in 2018 (Red: NIR, Green: Red Edge, Blue: Red).

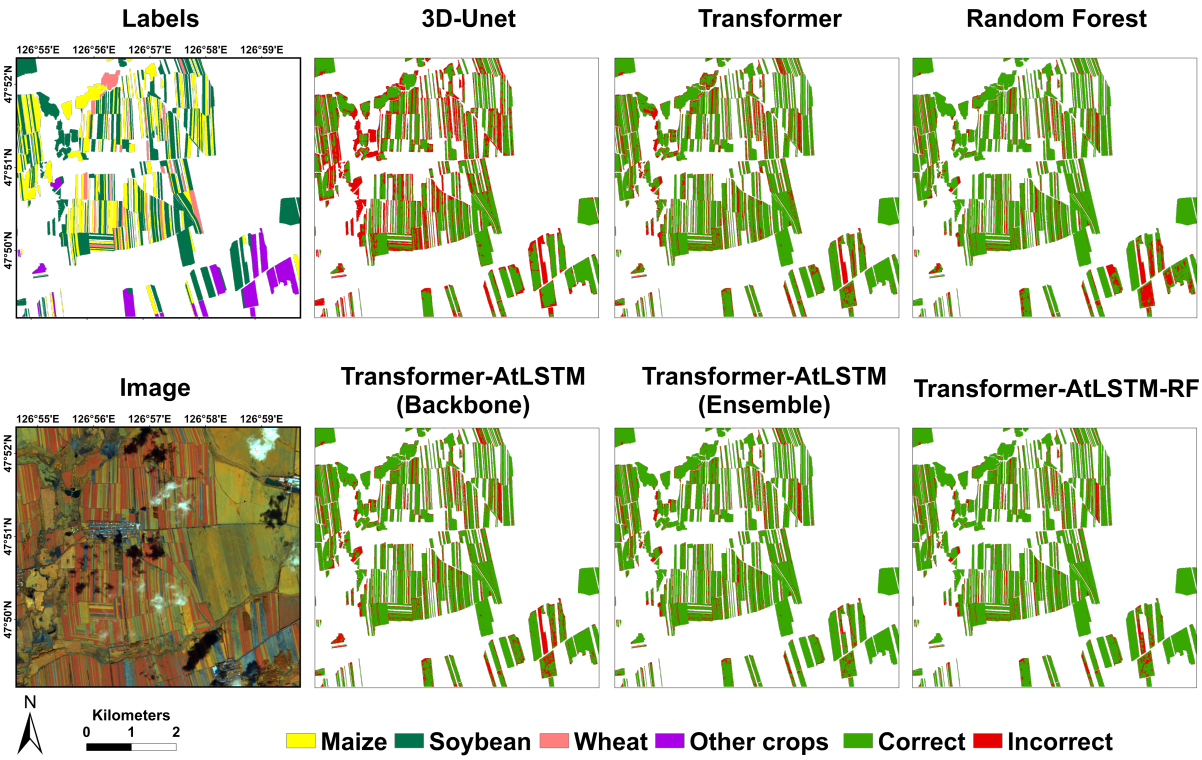

Figure 5-10. Crop mapping results for Site B in 2018. Captions follow Figure 5-9.

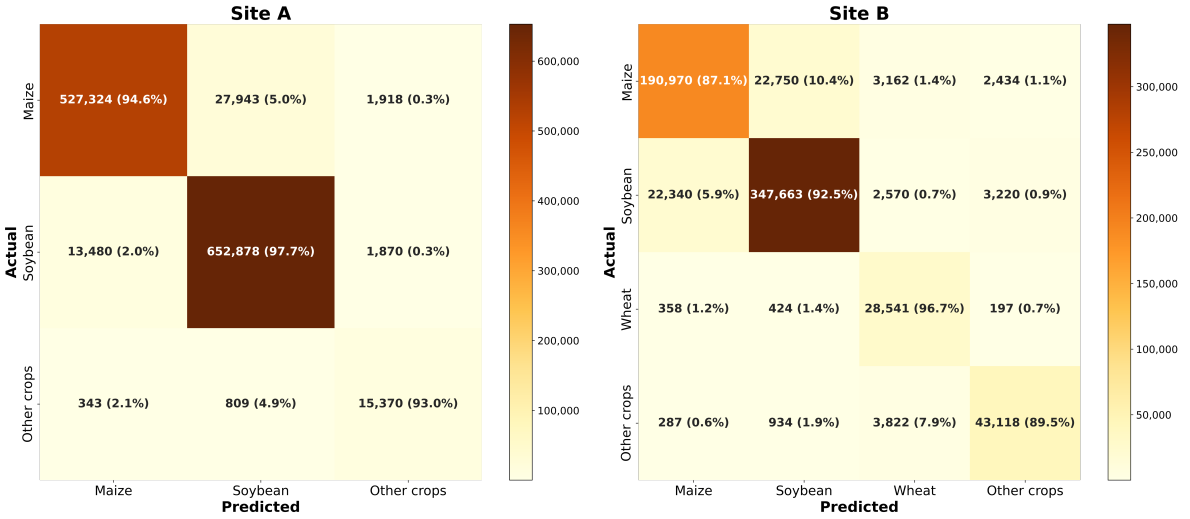

Figure 5-11. The confusion matrices of Site A and Site B by Transformer-AtLSTM-RF. Values in grids represent the number of samples along with their proportion (within brackets) calculated from each row. Main diagonal values stand for the number of correctly classified samples and percentages.

5.4.3 Evaluation of input feature importance

Feature importance analysis based on gradient backpropagation highlights the prominent features from two distinct components of the Transformer-AtLSTM (Ensemble) model. The analysis from each branch is visually presented in Figure 5-12 for AtLSTM and Figure 5-13 for Transformer (self-attention). Each displays the average gradients of attention weights with respect to the input data, effectively visualizing the temporal impact of different features and illustrating how certain features are emphasized during the model's prediction process. The AtLSTM model effectively combines the memory capabilities of bidirectional LSTM units with the selective focus provided by attention mechanisms, which provides a weighted contribution of input features at each time step. In the analysis, the VH and VV bands from different tracks exhibited similar trends of importance across all crop types. The model generally identified specific periods, particularly '2208-0309' and '0309-1509' in the VH and VV bands on Track 105, as important for all crop types in terms of the peak gradient distributions approaching 1. In track 32, the AtLSTM focused primarily on ‘1708-2908’ from the VV band, while all VH bands exhibit negative values for all crop types, suggesting negative influences on the classification for the current class. The lower importance values in track 32 correspond to its higher coherence compared to track 105 in the temporal profile in 2018 (Figure 5-8). Regarding optical bands, positive correlations between input features and crops were noted around '0809' in the NIR and Red bands from RapidEye for maize and soybean. Additionally, the model relied on bands B8a and B4 from Sentinel-2 on '0109' for wheat and soybean, and band B11 was more influential around '1908' and '1109' for maize, soybean, and wheat.

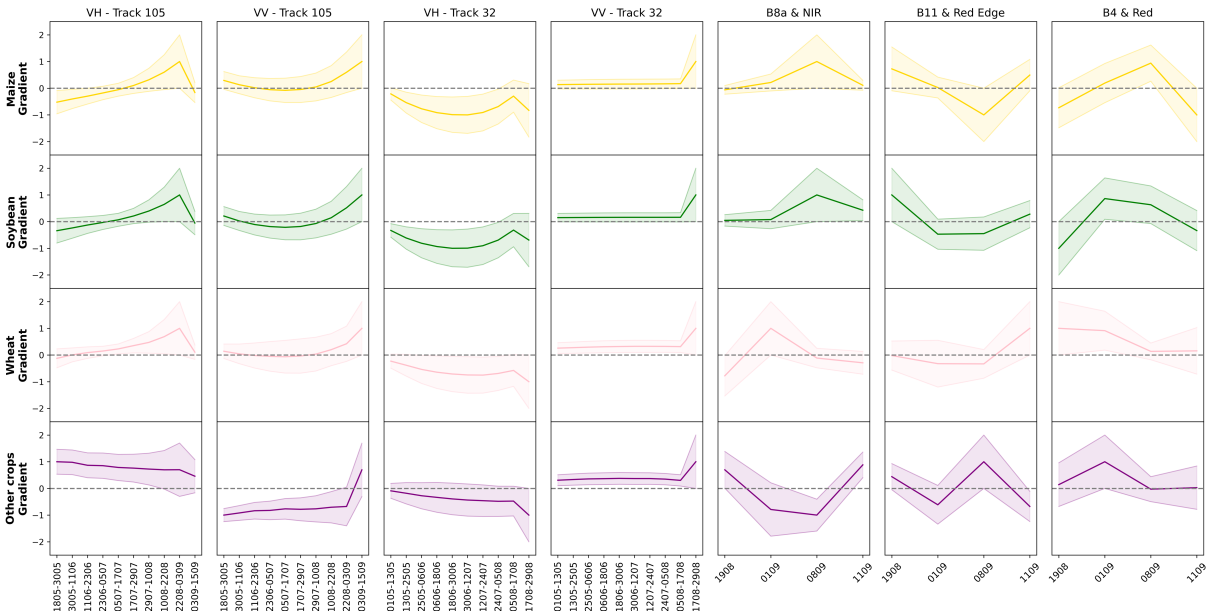

Figure 5-12. Average gradients of attention weights with respect to inputs from the AtLSTM end. 3000 samples of Site A and B were randomly selected from the attention weight layer for visualisation. The light buffer areas indicate one standard deviation from the average value. Positive (negative) values indicate a positive (negative) correlation between the predicted score and input features and vice versa. The value range was scaled from -1 to 1.

The multi-head attention mechanism in the Transformer model is designed to calculate self-attention weights for sequential features, effectively highlighting the most influential segments of the time series data. In Figure 5-13, all crop categories show uniform temporal importance (positive gradients) from both VH and VV bands in track 32, compared to the same track in AtLSTM (Figure 5-12). This observed uniformity of high importance across all crop types in track 32 aligns with its coherence trend, which is generally higher compared to track 105. Conversely, the consistent negative values across all crop types in track 105 correlate with a lower coherence pattern, as depicted in the temporal profile (Figure 5-8). In contrast, AtLSTM identified distinct peaks for all crops from track 105 (Figure 5-12). For certain optical bands, such as 'B8a & NIR', the Transformer model focused on specific dates ('0109' for maize and soybean, and '1908' for wheat), which were not prioritized by the AtLSTM. Conversely, the date '0809', which AtLSTM found significant for these crops in these bands, was overlooked by the Transformer. This difference in attention to specific time points of crops indicates that the attention behaviours of the Transformer and AtLSTM models could complement each other in the model's overall predictive process. When comparing these two models, similar patterns emerge at certain time points, such as '0809' for wheat from 'B11 & Red Edge', '0109' for maize, and '0809' for wheat and other crops from 'B4 & Red'. This suggests a level of consistency in the importance attributed to these specific periods by both models, reinforcing the value of

these time points in crop classification. This complementary effect when comparing these two models can also refer to ‘0809’ for wheat from ‘B11 & Red Edge’, ‘0109’ for maize, and ‘0809’ for wheat and other crops from ‘B4 & Red’.

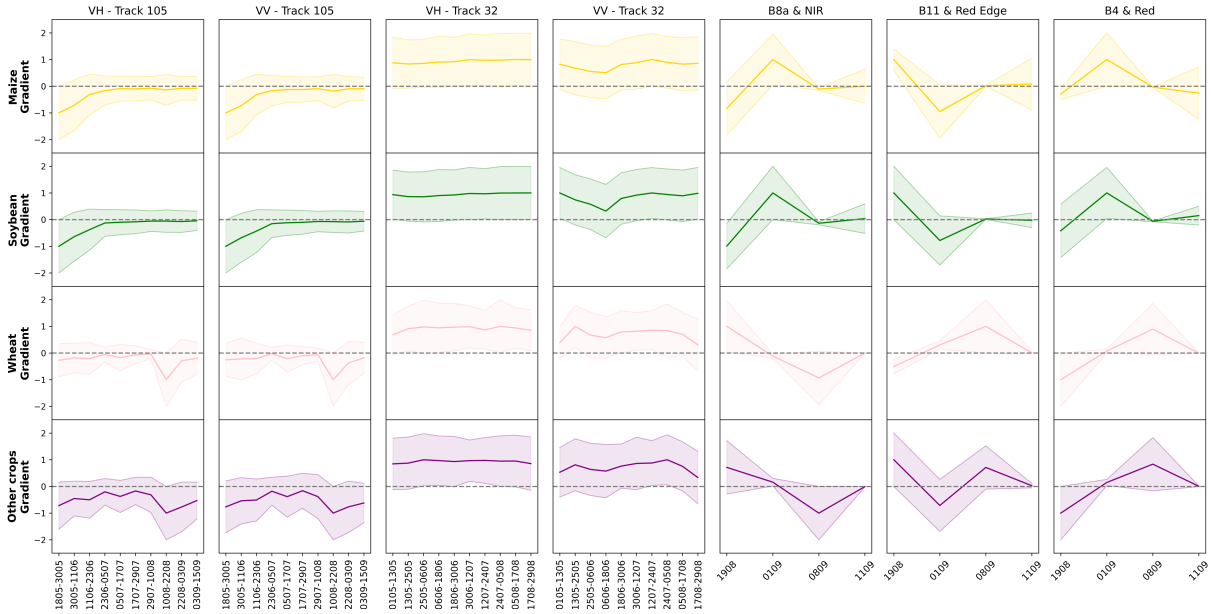

Figure 5-13. Average gradients of attention weights with respect to inputs from the Transformer end. 3000 samples of Site A and B were randomly selected from the second self-attention layer for visualisation. The light buffer areas indicate one standard deviation from the average value. Positive (negative) values indicate a positive (negative) correlation between the predicted score and input features and vice versa. The value range was scaled from -1 to 1.

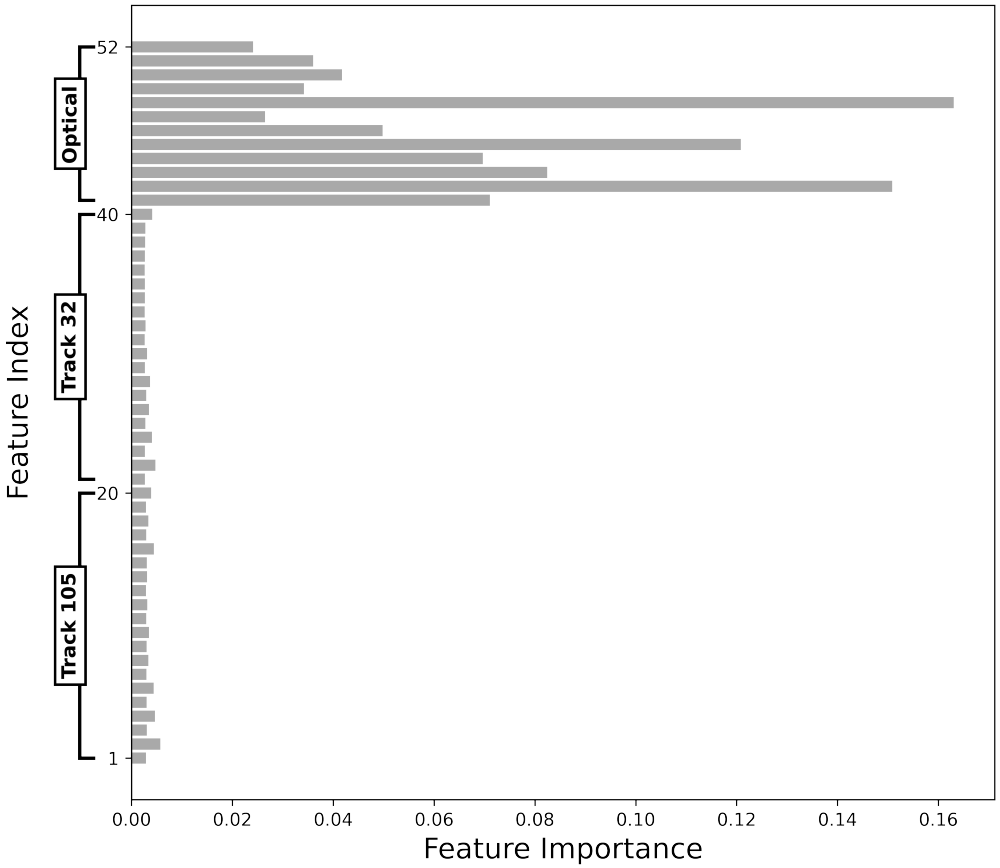

Figure 5-14. Feature importance values derived from the RF model, which sums up to 1. Track 105 and track 32 display a repeated sequence of VH and VV bands across their respective time steps. Track 105 has feature indices ranging from 1 to 20, corresponding to ten time steps (2 bands per time step). Track 32 covers indices 21 to 40. The 'Optical' data encompasses 12 features, out of which 9 are from Sentinel-2 (spanning indices 41 to 46, and 50 to 52) and 3 are from RapidEye (indices 47 to 49). Bands in Sentinel-2 for a single acquisition contain B8a, B11 and B4 in order, and RapidEye has NIR, Red Edge and Red.

5.5 Discussion

5.5.1 Performance analysis

5.5.1.1 Evaluation of model transferability

This study explored the potential of the combined use of InSAR coherence and multispectral bands in crop mapping with various deep learning models as well as an RF classifier as a baseline. The performance of models trained with the 2017 dataset was documented in Tables S1 and Table S2 and all models were trained and tested based on the block partitions in tiles in Figure 5-1. Subsequently, pre-trained models, including 3D U-Net, Transformer, and RF, were evaluated for their spatial and interannual transfer learning generalisability, particularly in regions with intercropping patterns. The transferability of these pre-trained models was found to be more effective for predominant crop classes, such as maize and soybean, compared to

less represented classes, like wheat and other crops, as demonstrated by the results obtained on both sites. Specifically, our study's RF outcomes contrasted with You and Dong (2020), who found that classifiers could be temporally transferred from the previous year with relative robustness due to the consistent crop phenology and spectral characteristics over short periods, which enables predictions without relying on current-year data. The discrepancy in findings could be attributed to differences in the scale of the study areas and the consistency of cropping patterns. Their research classified crops at the province level (Heilongjiang), possibly using a pre-trained RF model applied to regions at a broader scale with more homogenous crop patterns. Such patterns might smooth over localised variations that are significant in smaller-scale farming areas. In contrast, our study sites (A and B) feature distinctive intercropping practices (strip cropping) not typical in larger areas. This uniqueness may lead to reduced model transfer performance, as the classifiers' robustness relies on stable and consistent signals obtained in various spatiotemporal dynamics. Additionally, intercropping introduces increased signal variability from crops, potentially diminishing the accuracy of transfer learning for a model trained on data from the previous year.

The transferability of the models in this study was significantly improved for all crops by employing the fine-tuning technique with the 2018 crop samples. This improvement aligns with the findings of Hao et al. (2020) and Nowakowski et al. (2021), which suggest that disparities between training and testing samples, due to varying crop growth conditions, can lead to temporal data mismatches, resulting in misclassifications in transfer learning scenarios. The crop training samples could inherently reflect intra-class variability resulting from different meteorological conditions, soil types, topographical features, and cropping practices between regions (Zipper et al., 2016; Lobell and Azzari, 2017). However, the ability of pre-trained models to generalise was found to be less effective in testing sites with high spatial heterogeneity, particularly those not represented in the training data (Xu et al., 2020). This highlights the importance of fine-tuning models with additional data that are representative of the local testing environment. Nevertheless, single models in this study, such as 3D U-Net, Transformer, and RF, as indicated in Tables 5-2 – 5-5, were not sufficiently robust for mapping crops with specific cropping patterns and imbalanced distribution. The 3D U-Net model, which showed the highest OA (88%) and mIoU (72.7%) in the Bei'an 2017 test set (See Tables S1 and S2), encountered significant classification issues during transfer learning. The poor performance was derived because, despite being pre-trained with the 2017 CDL, it was fine-tuned using discretely distributed crop polygons within the ground-truth set instead of a

complete CDL mask for 2018. This led to increased learning of background values, and even setting a threshold of 70% for 2018 (See Section 5.3.3) could not mitigate this issue. Consequently, this research employed an ensemble approach, combining multiple classifiers with a rule-based strategy, namely Transformer-AtLSTM-RF, that improved overall and class-wise accuracy in crop mapping during transfer learning. This methodology resonates with similar strategies employed in other studies (e.g. Dou et al., 2021; Zhang et al., 2018), highlighting the effectiveness of ensemble models and rule-based decision fusion in enhancing the accuracy of crop classification in varied agricultural contexts.

5.5.1.2 Understanding of feature importance for crop mapping

Considering the potential variations in the environmental conditions and crop phenological characteristics between the training year (2017) and the test year (2018), our study analysed the impact of spatial and interannual variability on model behaviours by interpreting the input feature importance of the proposed model. This approach helps investigate specific temporal features that are most informative for the model, thereby enhancing the understanding of how the model adapts to and performs under certain agricultural applications.

In this study, gradient calculations with respect to input features were conducted separately for AtLSTM (Figure 12) and Transformer (Figure 13) within the Transformer-AtLSTM (Ensemble) framework. When analyzing InSAR coherence data over time, the AtLSTM component appeared to prioritize specific periods, particularly during the late growth and harvesting stages, to distinguish between different crops. This is evidenced by the relatively smooth distribution curves of feature importance across crop types. The AtLSTM model adopts an RNN structure to extract sequential relationships, likely contributing to its smooth feature importance curve. Specifically, the AtLSTM model transfers output features from one time point to the next in its bidirectional data transformation and aggregation pipelines, creating a sequence of dependencies. As a result, the feature importance of each current time point selectively aggregated information from the previous time period during the gradient backpropagation process (Xu et al., 2021b).

However, the Transformer model employs a self-attention mechanism to compute dependencies between all pairs of time positions in the input sequences, as opposed to using

recurrent structures. It does not sequentially accumulate information but rather extracts temporal dependencies between any two steps through its multi-head self-attention mechanism. Consequently, the Transformer's feature importance distribution, while not as smooth as AtLSTM's, reveals attention patterns that complement those derived by AtLSTM for improved crop mapping performance. This complementary relationship is consistent with the results presented in Tables 5-2 – 5-5. Overall, the combined use of VV and VH channels of coherence from the models yields accurate crop classification results. This effectiveness benefits from the complementary information provided by both polarizations, in agreement with findings from Mestre-Quereda et al. (2020).

In the analysis of gradient distributions for optical bands, the Transformer-AtLSTM (Ensemble) model pinpointed Sentinel-2's band 'B11 (SWIR)' from the date '1908' as a key feature for identifying maize and soybean. This finding agrees with several studies (e.g. Ghulam et al., 2008; Cai et al., 2018; Xu et al., 2021b; Wen et al., 2022), which have confirmed the importance of this band in differentiating between these two crops. Additionally, the model highlighted the importance of Sentinel-2's band 'B8a (Vegetation Red Edge)' from '0109' for maize and soybean, corroborating the findings of You and Dong (2020) regarding the effectiveness of red-edge bands in distinguishing these crops. In contrast, the 'Red Edge' band from RapidEye from '0809' did not appear as significant for maize and soybean in this study. This could be attributed to the fact that this study's single acquisition of RapidEye data might not have coincided with a period of pronounced spectral contrast between maize and soybean in the red-edge spectrum. Despite this, the model still recognised the importance of the bands 'B8a & NIR' and 'B4 & Red' for all crop types across both dates, '0109' and '0809'.

On the other hand, RF is highly dependent on multispectral features across the time series according to obtained importance scores (See Figure 5-14). When trained with 2018 data, RF outperformed the fine-tuned Transformer and 3D U-Net models across all crop types in terms of OA (95.1% in SiteA, 87.9% in Site B), mean F1 (90.8% in Site A, 80.8% in Site B) and mIoU (83.7% in Site A, 68.6% in Site B). These findings differ from previous studies (e.g. Xu et al., 2020; Rußwurm and Körner, 2020) that demonstrated Transformer was more effective than RF in crop mapping. However, those studies primarily utilised optical data with long time series, whereas the current study combined SAR-derived features with limited optical data. The restricted temporal coverage in our dataset may have limited the Transformer's ability to effectively extract long-sequence dependency features, as suggested by Shao and Bi (2022). In

contrast, RF can still perform well with a dataset that has limited sequential dependencies, making it particularly effective in scenarios with limited temporal depth of spectral features. Thus, the use of RF in a fusion of multitemporal SAR and optical data has proven to be an effective solution for accurate crop mapping, as also reported by Adrian et al. (2021).

5.5.2 Uncertainty and implication of transfer learning

This study investigated classifiers' capabilities to generalise new, previously unseen data from different geographical areas across a one-year temporal span. When applying transfer learning directly in Sites A and B of 2018, the models 3D U-Net, Transformer, and RF significantly underperformed in crop classification, leading to uncertainties about the impact of variations in phenological stages and crop calendars across different regions and years. The lower accuracy rates observed for both major and minor crops can be attributed to differences in crop growth conditions between the training areas and the transfer sites. Such mismatches often result from changing cropping practices, such as crop rotation and intercropping. These uncertainties can lead to poor performance in crop classification due to the significant differences between the ground-truth samples used in the training and testing regions (Wang et al., 2019; Hao et al., 2020). It is important to note that the same crop type may exhibit different spectral signatures when observed in distant regions or the same region over different times (Pohjankukka et al., 2017). Consequently, the mismatch in time-series data between training and testing regions necessitates the implementation of transfer learning techniques. Fine-tuning with new data can effectively adapt pre-trained models to the specific agricultural conditions and cropping patterns of the test regions. This approach helps to align the models more closely with the unique environmental and phenological characteristics of the new datasets, thereby improving the accuracy and reliability of the crop classification results.

While fine-tuning, ensemble learning, and rule-based strategies, and the combined use in the Transformer-AtLSTM-RF model, demonstrated promising potential in identifying crops through transfer learning, challenges persisted in accurately classifying the borders of crop fields. Identifying specific crops in intercropping areas was more challenging than in larger and homogenous croplands. This difficulty may be due to a mixture of surrounding cropland signals near field edges (Van Tricht et al., 2018). Another factor could be inaccuracies in the field labels at the borders or the limitations posed by the spatial resolution of the remote sensing data (Turkoglu et al., 2021). Despite the integration of fused SAR-related and multispectral

bands in this study, inaccuracies in classifying border pixels remained. However, both Site A and Site B suffered from sub-field contaminations from clouds and shadows in the RapidEye scenes (See Figure 5-9 and Figure 5-10). The incorporation of additional Sentinel-1 and Sentinel-2 features helped to overcome this occluded information to a considerable extent. The use of these complementary data sources synergistically enhances crop classification accuracy (e.g. Veloso et al., 2017; Van Tricht et al., 2018; Adrian et al., 2021; Blickensdörfer et al., 2022).

Future research could focus on using higher spatial resolution data to deal with those uncertainties and ameliorate the current mapping results. This approach would likely reduce misclassifications at the edges of small and narrow croplands and mitigate mixed-pixel impacts on mapping smallholder farms. Additionally, dense image time series could also help to identify consistent patterns of crop growth over time, thereby increasing classification accuracy despite pixel mixture at field borders. For instance, Mestre-Quereda et al. (2020) demonstrated improved crop classification using Sentinel-1 coherence with a 6-day interval compared to a 12-day temporal baseline. Additionally, compiling historical data over more years can enhance crop classification and model transferability (Cai et al., 2018; Xu et al., 2020). However, predicting crop types across different years remains challenging, especially for fields deviating from expected rotation patterns due to uncertain factors like changing farming practices, climate variability, natural hazards, resource shortages, soil degradation, policy shifts, and other socioeconomic dynamics (Zhang et al., 2021). To build a robust predictive model, the seasonal training set should include a wide variety of local samples that reflect diverse and representative crop sequence characteristics, with a focus on distinct intercropping patterns. Furthermore, integrating additional dynamic features that capture the temporal and spatial variability of intercropping, such as environmental parameters (Blickensdörfer et al., 2022) and meteorological variables (Zhong et al., 2014; Defourny et al., 2019), may further improve models' predictive performance.

5.6 Conclusion

This study demonstrated the enhanced performance of integrating InSAR coherence and multispectral bands for crop mapping using both deep learning and machine learning models. The significant potential of the proposed ensemble model learning architecture, combined with a rule-based decision fusion technique (Transformer-AtLSTM-RF), contributed to accurate

crop mapping for smallholder croplands with complex intercropping systems in Bei'an, China, in terms of the best quantitative and qualitative results overall compared to standalone models. The presented findings also highlighted the challenges of spatiotemporal transfer learning and the importance of fine-tuning pre-trained classifiers with representative data. The evaluation of input feature importances displayed the benefits of using multi-source data and the multi-model architecture, which contributed to complementary learning outcomes and identified key temporal features relevant to crop identification. This framework could lead to a more robust model capable of handling various scenarios. However, misclassifications around the borders of crop fields within intercropping systems remain a significant challenge. Future studies could explore using higher spatial resolution data, dense image time series, environmental variables, and collecting samples representative of specific agricultural conditions to ameliorate these limitations, combined with strategies similar to the developed classification framework.

References

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

Bargiel, D., 2017. A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sensing of Environment*, *198*, pp.369-383.

Bigdeli, B., Pahlavani, P. and Amirkolaee, H.A., 2021. An ensemble deep learning method as data fusion system for remote sensing multisensor classification. *Applied Soft Computing*, *110*, p.107563.

Blickensdörfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasmi, S. and 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*, p.112831.

Boyabatlı, O., Nasiry, J. and Zhou, Y., 2019. Crop planning in sustainable agriculture: Dynamic farmland allocation in the presence of crop rotation benefits. *Management Science*, *65*(5), pp.2060-2076.

Breiman, L., 2001. Random forests. *Machine learning*, *45*, pp.5-32.

Blaes, X. and Defourny, P., 2003. Retrieving crop parameters based on tandem ERS 1/2 interferometric coherence images. *Remote Sensing of Environment*, *88*(4), pp.374-385.

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B. and Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sensing of Environment*, *210*, pp.35-47.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. and Ronneberger, O., 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference*, *Athens*, *Greece*, *October 17-21*, *2016*, *Proceedings*, *Part II 19*, pp. 424-432.

Cioloş, D., Piebalgs, A., 2012. Sustainable agriculture for the future we want. *European Commission*, *Agriculture and Rural Development*.

Clinton, N., Yu, L. and Gong, P., 2015. Geographic stacking: Decision fusion to increase global land cover map accuracy. *ISPRS Journal of Photogrammetry and Remote Sensing*, *103*, pp.57-65.

Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaute, T. and Savinaud, M., 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*, pp.551-568.

Delrue, J., Bydekerke, L., Eerens, H., Gilliams, S., Piccard, I. and Swinnen, E., 2013. Crop mapping in countries with small-scale farming: A case study for West Shewa, Ethiopia. *International Journal of Remote Sensing*, *34*(7), pp.2566-2582.

Donezar, U., De Blas, T., Larrañaga, A., Ros, F., Albizua, L., Steel, A. and Broglia, M., 2019. Applicability of the multitemporal coherence approach to sentinel-1 for the detection and delineation of burnt areas in the context of the copernicus emergency management service. *Remote Sensing*, *11*(22), p.2607.

Dou, P., Shen, H., Li, Z. and Guan, X., 2021. Time series remote sensing image classification framework using combination of deep learning and multiple classifiers system. *International Journal of Applied Earth Observation and Geoinformation*, *103*, p.102477.

Du, P., Xia, J., Zhang, W., Tan, K., Liu, Y. and Liu, S., 2012. Multiple classifier system for remote sensing image classification: A review. *Sensors*, *12*(4), pp.4764-4792.

Ferretti, A., Monti-Guarnieri, A., Prati, C., Rocca, F. and Massonet, D., 2007. *InSAR Principles-guidelines for SAR Interferometry Processing and Interpretation*, *19*.

Gallo, I., Ranghetti, L., Landro, N., La Grassa, R. and Boschetti, M., 2023. In-season and dynamic crop mapping using 3D convolution neural networks and sentinel-2 time series. *ISPRS Journal of Photogrammetry and Remote Sensing*, *195*, pp.335-352.

Ghulam, A., Li, Z.L., Qin, Q., Yimit, H. and Wang, J., 2008. Estimating crop water stress with ETM+ NIR and SWIR data. *Agricultural and Forest Meteorology*, *148*(11), pp.1679-1695.

Hao, P., Di, L., Zhang, C. and Guo, L., 2020. Transfer Learning for Crop classification with Cropland Data Layer data (CDL) as training samples. *Science of The Total Environment*, *733*, p.138869.

Heihe Social and Economic Statistics Yearbook., 2018. *Heihe Social and Economic Statistics Yearbook*. Beijing: China Statistical Publishing House.

Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural Computation*, *9(8)*, pp.1735-1780.

Hsu, C.W., Chang, C.C. and Lin, C.J., 2003. A practical guide to support vector classification, pp. 1396-1400.

Jia, K., Wu, B. and Li, Q., 2013. Crop classification using HJ satellite multispectral data in the North China Plain. *Journal of Applied Remote Sensing*, *7*(1), pp.073576-073576.

Juventia, S.D., Norén, I.L.S., Van Apeldoorn, D.F., Ditzler, L. and Rossing, W.A., 2022. Spatio-temporal design of strip cropping systems. *Agricultural Systems*, *201*, p.103455.

King, L., Adusei, B., Stehman, S.V., Potapov, P.V., Song, X.P., Krylov, A., Di Bella, C., Loveland, T.R., Johnson, D.M. and Hansen, M.C., 2017. A multi-resolution approach to national-scale cultivated area estimation of soybean. *Remote Sensing of Environment*, *195*, pp.13-29.

Konduri, V.S., Kumar, J., Hargrove, W.W., Hoffman, F.M. and Ganguly, A.R., 2020. Mapping crops within the growing season across the United States. *Remote Sensing of Environment*, *251*, p.112048.

Lechner, A.M., Stein, A., Jones, S.D. and Ferwerda, J.G., 2009. Remote sensing of small and linear features: Quantifying the effects of patch size and length, grid position and detectability on land cover mapping. *Remote Sensing of Environment*, *113*(10), pp.2194-2204.

Li, C., Hoffland, E., Kuyper, T.W., Yu, Y., Zhang, C., Li, H., Zhang, F. and van der Werf, W., 2020. Syndromes of production in intercropping impact yield gains. *Nature Plants*, *6*(6), pp.653-660.

Li, H., Zhang, C., Zhang, S. and Atkinson, P.M., 2019. A hybrid OSVM-OCNN method for crop classification from fine spatial resolution remotely sensed imagery. *Remote Sensing*, *11*(20), p.2370.

Li, R., Xu, M., Chen, Z., Gao, B., Cai, J., Shen, F., He, X., Zhuang, Y. and Chen, D., 2021. Phenology-based classification of crop species and rotation types using fused MODIS and Landsat data: The comparison of a random-forest-based model and a decision-rule-based model. *Soil and Tillage Research*, *206*, p.104838.

Liu, N., Zhao, Q., Williams, R. and Barrett, B., 2023. Enhanced crop classification through integrated optical and SAR data: a deep learning approach for multi-source image fusion. *International Journal of Remote Sensing*, pp.1-29.

Livingston, M., Roberts, M.J. and Zhang, Y., 2015. Optimal sequential plantings of corn and soybeans under price uncertainty. *American Journal of Agricultural Economics*, *97*(3), pp.855-878.

Lobell, D.B. and Azzari, G., 2017. Satellite detection of rising maize yield heterogeneity in the US Midwest. *Environmental Research Letters*, *12*(1), p.014014.

Ma, X., Fu, A., Wang, J., Wang, H. and Yin, B., 2018. Hyperspectral image classification based on deep deconvolution network with skip architecture. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(8), pp.4781-4791.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U. and Gascon, F., 2017, October. Sen2Cor for Sentinel-2. In *Image and Signal Processing for Remote Sensing XXIII*, *10427*, pp. 37-48.

Manavalan, R., 2018. Review of synthetic aperture radar frequency, polarization, and incidence angle data for mapping the inundated regions. *Journal of Applied Remote Sensing*, *12*(2), pp.021501-021501.

Mandal, D., Kumar, V., Ratha, D., Dey, S., Bhattacharya, A., Lopez-Sanchez, J.M., McNairn, H. and Rao, Y.S., 2020. Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data. *Remote Sensing of Environment*, *247*, p.111954.

Mazarire, T.T., Ratshiedana, P.E., Nyamugama, A., Adam, E. and Chirima, G., 2020. Exploring machine learning algorithms for mapping crop types in a heterogeneous agriculture landscape using Sentinel-2 data. A case study of Free State Province, South Africa. *South African Journal of Geomatics*, *9*(2), pp.333-347.

Mestre-Quereda, A., Lopez-Sanchez, J.M., Vicente-Guijalba, F., Jacob, A.W. and Engdahl, M.E., 2020. Time-series of Sentinel-1 interferometric coherence and backscatter for crop-type mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, pp.4070-4084.

Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Gill, E. and Molinier, M., 2019. A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem. *ISPRS Journal of Photogrammetry and Remote Sensing*, *151*, pp.223-236.

Nasirzadehdizaji, R., Cakir, Z., Sanli, F.B., Abdikan, S., Pepe, A. and Calo, F., 2021. Sentinel-1 interferometric coherence and backscattering analysis for crop monitoring. *Computers and Electronics in Agriculture*, *185*, p.106118.

Nowakowski, A., Mrziglod, J., Spiller, D., Bonifacio, R., Ferrari, I., Mathieu, P.P., Garcia-Herranz, M. and Kim, D.H., 2021. Crop type mapping by using transfer learning. *International Journal of Applied Earth Observation and Geoinformation*, *98*, p.102313.

Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B. and Glocker, B., 2018. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

Onojeghuo, A.O., Miao, Y. and Blackburn, G.A., 2023. Deep ResU-Net Convolutional Neural Networks Segmentation for Smallholder Paddy Rice Mapping Using Sentinel 1 SAR and Sentinel 2 Optical Imagery. *Remote Sensing*, *15*(6), p.1517.

Phalke, A.R. and Özdoğan, M., 2018. Large area cropland extent mapping with Landsat data and a generalized classifier. *Remote Sensing of Environment*, *219*, pp.180-195.

Rußwurm, M. and Körner, M., 2018. Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS International Journal of Geo-Information*, *7*(4), p.129.

Rußwurm, M. and Körner, M., 2020. Self-attention for raw optical satellite time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *169*, pp.421-435.

Sahajpal, R., Zhang, X., Izaurralde, R.C., Gelfand, I. and Hurtt, G.C., 2014. Identifying representative crop rotation patterns and grassland loss in the US Western Corn Belt. *Computers and Electronics in Agriculture*, *108*, pp.173-182.

Saini, R. and Ghosh, S.K., 2018. Crop classification on single date sentinel-2 imagery using random forest and suppor vector machine. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, pp.683-688.

Shao, R. and Bi, X.J., 2022. Transformers meet small datasets. *IEEE Access*, *10*, pp.118454-118464.

Shendryk, Y., 2019. Deep learning for multi-modal classification of cloud, shadow and land cover scenes in PlanetScope and Sentinel-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *13*.

Steele-Dunne, S.C., McNairn, H., Monsivais-Huertero, A., Judge, J., Liu, P.W. and Papathanassiou, K., 2017. Radar remote sensing of agricultural canopies: A review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(5), pp.2249-2273.

Sica, F., Pulella, A., Nannini, M., Pinheiro, M. and Rizzoli, P., 2019. Repeat-pass SAR interferometry for land cover classification: A methodology using Sentinel-1 Short-Time-Series. *Remote Sensing of Environment*, *232*, p.111277.

Tao, C., Meng, Y., Li, J., Yang, B., Hu, F., Li, Y., Cui, C. and Zhang, W., 2022. MSNet: multispectral semantic segmentation network for remote sensing images. *GIScience & Remote Sensing*, *59*(1), pp.1177-1198.

Teluguntla, P., Thenkabail, P.S., Oliphant, A., Xiong, J., Gumma, M.K., Congalton, R.G., Yadav, K. and Huete, A., 2018. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS Journal of Photogrammetry and Remote Sensing*, *144*, pp.325-340.

Touzi, R., Lopes, A., Bruniquel, J. and Vachon, P.W., 1999. Coherence estimation for SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(1), pp.135-149.

Turkoglu, M.O., D'Aronco, S., Perich, G., Liebisch, F., Streit, C., Schindler, K. and Wegner, J.D., 2021. Crop mapping from image time series: Deep learning with multi-scale label hierarchies. *Remote Sensing of Environment*, *264*, p.112603.

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

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

Volpi, M. and Tuia, D., 2016. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(2), pp.881-893.

Wang, S., Azzari, G. and Lobell, D.B., 2019. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment*, *222*, pp.303-317.

Wang, S., Yang, Y., Luo, Y. and Rivera, A., 2013. Spatial and seasonal variations in evapotranspiration over Canada's landmass. *Hydrology and Earth System Sciences*, *17*(9), pp.3561-3575.

Wei, P., Chai, D., Lin, T., Tang, C., Du, M. and Huang, J., 2021. Large-scale rice mapping under different years based on time-series Sentinel-1 images using deep semantic segmentation model. *ISPRS Journal of Photogrammetry and Remote Sensing*, *174*, pp.198-214.

Wei, S., Zhang, H., Wang, C., Wang, Y. and Xu, L., 2019. Multi-temporal SAR data large-scale crop mapping based on U-Net model. *Remote Sensing*, *11*(1), p.68.

Wen, Y., Li, X., Mu, H., Zhong, L., Chen, H., Zeng, Y., Miao, S., Su, W., Gong, P., Li, B. and Huang, J., 2022. Mapping corn dynamics using limited but representative samples with adaptive strategies. *ISPRS Journal of Photogrammetry and Remote Sensing*, *190*, pp.252-266.

Wu, X., Xiao, X., Yang, Z., Wang, J., Steiner, J. and Bajgain, R., 2021. Spatial-temporal dynamics of maize and soybean planted area, harvested area, gross primary production, and grain production in the Contiguous United States during 2008-2018. *Agricultural and Forest Meteorology*, *297*, p.108240.

Xia, T., He, Z., Cai, Z., Wang, C., Wang, W., Wang, J., Hu, Q. and Song, Q., 2022. Exploring the potential of Chinese GF-6 images for crop mapping in regions with complex agricultural landscapes. *International Journal of Applied Earth Observation and Geoinformation*, *107*, p.102702.

Xu, J., Yang, J., Xiong, X., Li, H., Huang, J., Ting, K.C., Ying, Y. and Lin, T., 2021b. Towards interpreting multi-temporal deep learning models in crop mapping. *Remote Sensing of Environment*, *264*, p.112599.

Xu, L., Zhang, H., Wang, C., Wei, S., Zhang, B., Wu, F. and Tang, Y., 2021a. Paddy rice mapping in thailand using time-series sentinel-1 data and deep learning model. *Remote Sensing*, *13*(19), p.3994.

You, N. and Dong, J., 2020. Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, *161*, pp.109-123.

You, N., Dong, J., Huang, J., Du, G., Zhang, G., He, Y., Yang, T., Di, Y. and Xiao, X., 2021. The 10-m crop type maps in Northeast China during 2017–2019. *Scientific Data*, *8*(1), p.41.

You, N., Dong, J., Li, J., Huang, J. and Jin, Z., 2023. Rapid early-season maize mapping without crop labels. *Remote Sensing of Environment*, *290*, p.113496.

Yuan, Y. and Lin, L., 2020. Self-supervised pretraining of transformers for satellite image time series classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, pp.474-487.

Zhang, C., Di, L., Hao, P., Yang, Z., Lin, L., Zhao, H. and Guo, L., 2021. Rapid in-season mapping of corn and soybeans using machine-learned trusted pixels from Cropland Data Layer. *International Journal of Applied Earth Observation and Geoinformation*, *102*, p.102374.

Zhang, C., Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J. and Atkinson, P.M., 2018. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *140*, pp.133-144.

Zhang, H., Wang, Y., Shang, J., Liu, M. and Li, Q., 2021. Investigating the impact of classification features and classifiers on crop mapping performance in heterogeneous agricultural landscapes. *International Journal of Applied Earth Observation and Geoinformation*, *102*, p.102388.

Zhang, C., Di, L., Lin, L., Li, H., Guo, L., Yang, Z., Eugene, G.Y., Di, Y. and Yang, A., 2022. Towards automation of in-season crop type mapping using spatiotemporal crop information and remote sensing data. *Agricultural Systems*, *201*, p.103462.

Zhong, L., Gong, P. and Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. *Remote Sensing of Environment*, *140*, pp.1-13.

Zhong, L., Hu, L. and Zhou, H., 2019. Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, *221*, pp.430-443.

Zhou, T., Pan, J., Zhang, P., Wei, S. and Han, T., 2017. Mapping winter wheat with multi-temporal SAR and optical images in an urban agricultural region. *Sensors*, *17*(6), p.1210.

Zipper, S.C., Qiu, J. and Kucharik, C.J., 2016. Drought effects on US maize and soybean production: spatiotemporal patterns and historical changes. *Environmental Research Letters*, *11*(9), p.094021.

Chapter 6 Discussion

Crop mapping across large areas using remote sensing imagery can be challenging. The primary issue is the often similar spectral and/or backscatter characteristics at various phenological stages throughout the crop growing season, which can result in misclassification. Additionally, the complex spatiotemporal relationships among crops can adversely affect the performance and transferability of classification algorithms, especially in terms of annual variations in cropping patterns observed in major grain-producing regions like Northeast China. Existing classification methods have not been fully explored for effectively analysing multi-temporal and multi-source remote sensing data, which necessitates the need for the development of advanced methodologies to overcome the limitations caused by increased inter-class similarity and intra-class heterogeneity in satellite-based crop mapping (Interdonato et al., 2019; Hu et al., 2019; You et al., 2021; You et al., 2023).

Deep learning has recently emerged as a transformative technology across various research domains including the field of remote sensing. Image classification, in particular, has significantly benefited from deep learning's ability to analyse intricate relationships within high-dimensional data and perform feature learning and automatic extraction in an end-to-end fashion without extensive domain knowledge and expertise (LeCun et al., 2015). Mirroring human visual cognition, deep learning algorithms hierarchically process images through multiple layers represented by multi-level features. This capability enables the applications of temporal or spatiotemporal modelling tasks, such as crop mapping and highlights the potential of deep learning in enhancing feature representations for identifying unique characteristics of different crops in time series remote sensing data.

Although deep learning negates the need for most manual feature engineering, the selection and fusion of input features derived from multi-source remote sensing data can decisively influence the model's classification performance (Zhong et al., 2019; Liao et al., 2020; Yang et al., 2020; Adrian et al., 2021; Teimouri et al., 2022). It is still important to determine the most informative features from remote sensing data for input fed into deep learning models in order to accurately differentiate between crop types across their growth stages. Furthermore, the physical interpretation of deep learning models is needed in terms of visualising and understanding the model's learning process in the context of crop mapping.

This thesis developed a set of classification frameworks based on deep learning architectures to produce accurate crop maps at multiple scales, using multi-temporal and multi-source satellite remote sensing data, specifically focusing on Bei'an in Northeast China. These crop classification frameworks evaluated SAR-derived features and the synergetic use of optical and SAR data using novel deep learning architectures that incorporate Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The research evaluated various ways of integration designed for different model learning mechanisms and then compared them with other popular models to validate their effectiveness in differentiating crops. Nonetheless, several inherent challenges and uncertainties persist in applying deep learning models for crop mapping. These include optimising model complexity, enhancing spatial and interannual model transferability, and addressing high misclassification rates near cropland boundaries. Additionally, this thesis explored the interpretability of the learning process inherent in deep neural networks, through several aspects. The major contributions of this research are as follows:

(1) Comprehensive evaluations of diverse input features for crop mapping derived from multi-temporal SAR and multispectral data.

(2) The design and implementation of novel deep learning architectures, in terms of various ensemble learning strategies adapted to accommodate the complexities of cropland distributions and cropping patterns.

(3) Multi-perspective interpretation of the deep learning networks through visualising the feature learning process, analysing soft outputs, and assessing the input feature importance by identifying the parts of the time series data influencing the model's decisions.

This chapter synthesises the research findings from each experimental chapter (Chapters 3 to 5) and discusses the limitations of the research, followed by recommendations for future research directions.

6.1 A Comparative Synthesis of Deep Learning Models in Crop Mapping

Chapter 3 focuses on pixel-based crop classification using temporal models, wherein it compares Conv1D (Zhong et al., 2019), Conv1D-RF (Yang et al., 2020), and Transformer (Rußwurm and Körner, 2020) models against the newly designed Conv1D-LSTM model. Zhong et al. (2019) developed the Conv1D model, incorporating inception modules that combine 1D-CNN and max-pooling layers of varying sizes to process multi-scale features in parallel. This design enhances feature extraction across different temporal dimensions, which is better for identifying various crops along their growth stages. Yang et al. (2020) modified this architecture by replacing the Softmax classifier with a Random Forest (RF) classifier, thereby leveraging the feature extraction capabilities of 1D-CNN and the decision-making of the tree-based classifier. Rußwurm and Körner (2020) employed a Transformer model built upon a multi-head self-attention mechanism, focusing on specific segments of the time series data. Each model achieved the highest overall accuracy (OA) of 88% and the second-best average F1 score of 83% for crops in a specific scenario. In contrast, the Conv1D-LSTM was constructed as ensemble learning using multi-depth 1D-CNN and attention-based LSTM. This architecture synergises the strengths of both networks in feature fusion (Interdonato et al., 2019; Hamad et al., 2020), demonstrating improved performance in managing inherently imbalanced class distribution in real-life datasets. Notably, it achieved the highest OA of 88%, and an average F1 score of 84% using only m-chi decomposition features as the time series inputs. Moreover, certain models, like the Transformer, suffered from computational and memory costs due to the quadratic increase in attention scores during the self-attention process (Xu et al., 2021). The Conv1D-RF also encountered similar issues when handling large training dataset dimensions, where the complexity of the forest's trees increased significantly, which in turn required more resources for optimal split determination and information storage. As presented in Chapter 3 (Figure S3), the Conv1D-LSTM model required comparatively less training time than other temporal models and was found to efficiently surpass other pixel-wise deep learning models in large-scale crop mapping using a minimal number of SAR-derived features. This finding aligns with previous studies (e.g. Hamad et al. 2020) that emphasize the effectiveness of applying joint ensemble learning architectures to address imbalanced datasets.

Chapter 4 transitions from relying on pixel-based classifications to adopting patch-based classifiers. Pixel-based classification, which predominantly focuses on extracting features along the temporal dimension, often results in the ‘salt-and-pepper’ effect in some land units,

often requiring post-classification techniques for smoothing noise (Zhang et al., 2021). Patch-based classifiers, in contrast, are designed to extract spatiotemporal dependencies within multi-temporal remote sensing data to control such noise effectively. This chapter performed a comparative analysis of various typical architectures for crop mapping, including 3D-CNNs (Ji et al., 2018), ConvSTAR (Turkoglu et al., 2021), the 1D-CNN-based TCNN model (Pelletier et al., 2019), and a 3D-2D CNN (Roy et al., 2020) and introduces the novel 3D-ConvSTAR model, which connects 3D-CNN, ConvSTAR, and a shallow 2D-CNN through a hybrid sequential processing approach. This model differs from the Conv1D-LSTM presented in Chapter 3, as the latter processes separate inputs for each module, and combines their resulting outputs, while 3D-ConvSTAR proceeds with a stacked time series input vector in interdependent learning processes across the different models. 3D-ConvSTAR demonstrated higher accuracy in crop mapping over standalone and combined architectures such as TCNN, 3D-CNN, and ConvSTAR. It achieved an OA of 91.7%, a Kappa coefficient of 85.7%, and a mean F1 score of 87.7%. While data augmentation techniques, notably the mix-up method, enhanced classification performance for specific crops like maize and soybean, they reduced performance for wheat and other crops. Oversampling, another augmentation method, proved generally effective but was insufficient in addressing imbalanced class distribution compared to the 3D-ConvSTAR model without data augmentation. However, it is important to note that the 3D-ConvSTAR network is prone to generating more training parameters, which leads to longer model training times. This is mostly attributed to the complexity of its architecture, which sequentially connects different networks for feature extraction from more than one dimension, thereby requiring more computational resources.

Chapter 5 shifts the focus from crop mapping at a county level, as experimented in Chapters 3 and 4, to assessing model transferability at localised scales. In this chapter, an innovative framework is introduced, extending ensemble learning of temporal models through a rule-based strategy to enhance the decision-making process in crop classification. This framework features a multi-stream architecture, combining Transformer and Attention-based LSTM (AtLSTM), to process three distinct time series inputs separately and in parallel. This design allows each stream to optimise its input processing before merging the learned features in subsequent stages. The outputs from each stream are concatenated and integrated with probabilities derived from an RF classifier, using predefined thresholds. This fusion of probabilities showcases the complementary strengths of self-attention, attention-based LSTM, and the tree-based mechanism of RF. This decision fusion framework, Transformer-AtLSTM-

RF, demonstrated effective spatial and interannual transfer learning capabilities for crop mapping across two local sites, as evidenced by the highest OA and mean Intersection over Union (mIoU) metrics (See Tables 5-2 – 5-5). Compared with the hybrid architecture, namely Transformer-AtLSTM (Backbone), standalone Transformer, and 3D U-Net models, this decision fusion framework outperformed. Considering the 3D U-Net employs fully connected networks (FCNs) for spatiotemporal feature extraction, its effectiveness was subject to the quality of ground-truth data during model fine-tuning.

Moreover, this decision-fusion technique aligns with similar approaches in other studies (e.g., Dou et al., 2021; Zhang et al., 2018), highlighting the effectiveness of ensemble models and rule-based decision fusion in enhancing crop classification accuracy across various agricultural settings. Additionally, fine-tuning pre-trained models with 2018 crop samples significantly improved transferability for all crop classes, consistent with research emphasising the importance of aligning training and testing samples with local agricultural conditions (Hao et al., 2020; Nowakowski et al., 2021). Nevertheless, the complex intercropping patterns in the transfer sites introduced increased signal variability and resulted in misclassifications around cropland boundaries during transfer learning. This limitation is not necessarily attributed to the design of the model architectures. It could potentially be mitigated by using higher spatial resolution (> 5m) and denser time series remote sensing data, coupled with training samples that mostly represent diverse intercropping patterns.

In summary, a comprehensive evaluation of deep learning networks for large-scale crop mapping is presented, covering both pixel-based (Chapter 3) and patch-based networks (Chapter 4), as well as assessing the transferability of designated models for local-scale areas with complex intercropping systems (Chapter 5). Each chapter explored the way of jointly constructing multi-branch model architectures to learn features effectively. Collectively, the experimental results contribute significantly to the applications of deep neural networks by offering designed architectures that enhance crop mapping and provide greater potential for broader applications.

6.2 Synergetic Use of SAR and Optical Data in Enhancing Crop Mapping

This thesis critically evaluates the role of input features derived from satellite remote sensing data, using feature selection techniques and deep learning models for crop mapping. In Chapter 3, a selection of SAR-derived features prevalent in crop mapping and growth monitoring was calculated from Sentinel-1 data. These included backscatter, GLCM features, radar-based vegetation indices, covariance parameters, and polarimetric decomposition features. Through feature selection and ablation experiments, m-chi decomposition features were identified as particularly efficient and effective in distinguishing crops on a large scale, which aligns with previous findings using machine learning models (Sonobe et al., 2019; Mahdianpari et al., 2019; Dingle et al., 2022). While the use of quad-polarimetric (quad-pol) data has advantages in crop mapping compared to dual-polarimetric (dual-pol) features (Xie et al., 2019; Liao et al., 2020; He et al., 2020), the application of satellite-based quad-pol sensors is limited by their swath coverage (Raney, 2019). This limitation constrains their practicality for large-scale crop map prediction. Moreover, accessing dense time-series quad-pol data can be financially prohibitive. Insufficient multi-temporal data could lead to inaccuracies in classification, given that the in-season classification scheme discussed in Chapter 3 demonstrated that model performance was initially marginal in early growth stages due to the inadequate acquisition of SAR input features.

Chapter 4 integrates SAR-derived features with optical-derived features and presents a significant progression in crop classification accuracy. Specifically, the experiment incorporated three Sentinel-2 acquisitions into twenty-three Sentinel-1 acquisitions. Optical data for each acquisition involve Red, Vegetation Red Edge, and SWIR bands, which effectively differentiate soybean and maize in Northeast China (You and Dou, 2020). The SAR data and multi-spectral bands were sequentially stacked according to the acquisition dates across the crops' growth stages, creating a fused vector as the model input. This approach, stacking SAR and optical data, contributes to improved classification performance compared to using multi-temporal SAR data alone. Particularly, models incorporating multi-spectral bands and polarimetric features (m-chi features) exhibited higher accuracy in all tested scenarios. For instance, the most effective model using the Optical+m-chi input scenario achieved improvements of over 2% in OA, more than 4% in Kappa, and over 6% in mean F1 score compared to only using m-chi features as inputs (Tables 4-1 – 4-2). Chapter 4 thus demonstrated that combining input features from SAR and optical sources into a compatible tensor for the proposed model effectively leveraged the synergy of SAR and optical data to

improve crop classification. Furthermore, this fusion of optimally selected features mitigated the risk of losing original information, which could occur in directly stacking and normalising the fused raw dataset into a common tensor.

Chapter 5 investigated the fusion of InSAR coherence and multi-spectral bands for crop classification. Since multi-temporal coherence was calculated based on image pairs of different orbital tracks of Sentinel-1, stacking all feature bands (as done in Chapter 4) would not maintain a consistent 12-day interval for coherence data over the timeframe. Consequently, the experiment processed the time series coherence based on tracks and optical bands in separate streams, with the resulting features from each stream being concatenated channel-wise. This multi-branch model architecture, similar to the Conv1D-LSTM proposed in Chapter 3, takes multi-source inputs and combines feature vectors generated from each branch. This approach allows each branch to contribute complementary information to the discriminative learning process, providing different perspectives (Interdonato et al., 2019; Hamad et al., 2020). Additionally, the 3D U-Net architecture in Chapter 5, inspired by Tao et al. (2022), adopts a dual-branch parallel U-Net architecture. This design fuses learned semantic features from SAR and optical sources at each level of the model's upsampling stages.

Comparing the fusion techniques for SAR and optical data is challenging in terms of feature fusion through stacking (Chapter 4) versus through model architecture (Chapter 5), due to differences in input features, model architectures, and testing areas. A comparative analysis using m-chi decomposition features and InSAR coherence was subsequently performed (See Figures S9 and S10). Both were calculated from Sentinel-1 and combined with Sentinel-2 and RapidEye multi-spectral bands. Modelling was conducted under identical experimental conditions and the same model implementation but with different inputs. The Transformer-AtLSTM (Ensemble), which outperformed its backbone version (Tables 5-2 – 5-5), was used for the comparison. The Transformer component of the model was pre-trained with 2017 data and fine-tuned with 2018 data. Results indicate that incorporating coherence with optical data marginally increased overall mapping performance by > 1% of OA, > 3% of mean F1 and > 5% of mIoU for both sites during model transfer learning (Tables S3 to S6). This finding demonstrates the potential of coherence as an effective input feature for crop classification, but a different model architecture might be useful to fully leverage polarimetric decomposition features like the one proposed in Chapter 4. Future experiments would be worth comparing these input features across various model architectures and conditions, including crop types,

testing areas, fusion strategies, and transfer learning. Additionally, evaluating the computational efficiency of these models is also important to ensure their robustness and generalisability.

In summary, this research has successfully evaluated a range of satellite-based features and how they are used synergistically to enhance crop mapping in Bei'an County. Chapter 3 focused on various SAR-derived features from Sentinel-1 and identified m-chi decomposition as the most effective input for deep learning models in crop mapping on a large scale. Chapter 4 advanced crop mapping by integrating SAR polarimetric features with a few multi-spectral data from Sentinel-2. Chapter 5 focused on mapping local croplands with complex cropping patterns, using multi-temporal InSAR coherence and high-resolution multi-spectral data.

6.3 Interpretation of Deep Learning Models in Crop Mapping

This research addressed challenges in interpreting deep learning models due to their 'black box' nature, where the intricate model training process can affect a comprehensive understanding of how inputs are transformed into outputs. It sets clear linkages between each experimental chapter to interpret the designed deep learning models for crop mapping. This interpretation involves hidden feature analysis in Chapter 3, gradient and soft output analysis in Chapter 4, and input feature importance evaluation in Chapter 5.

In Chapter 3, the Conv1D-LSTM model combines multi-level Conv1D with attention-based LSTM networks to capture temporal dynamics and features for crop classification. From the CNN end, the weight distribution across the timeframe for different classes is visualised over multi-level Conv1D layers (Figures 3-14 – 3-15). The fluctuations in growing patterns show how the model learns to prioritize pixels in the time series. Starting with the shallowest layer, the weights are estimated to be small-scale variations, with a sparse distribution around August. When the input series passes through the intermediate layers and the deepest layers, the weights start to disperse and become scattered, becoming localised at specific acquisition dates. This hierarchical aggregation of simple weight distribution patterns into more complex ones aligns with Zhong et al. (2019), who used guided back-propagation to identify the most activated parts of the input series in crop classification. Furthermore, the attention mechanism in the LSTM illustrates how cumulative sequential data becomes increasingly significant for identifying crops (Figure 3-16). This is also evident in the in-season classification using t-

distributed stochastic neighbour embedding (t-SNE) for high-dimensional feature representation (Van der Maaten and Hinton, 2008), which shows crop types becoming more distinctly segregated in feature space as the model learns through accumulated phenological features (Figure 3-17). However, the challenge persists in differentiating crops using dual-polarized SAR data as model inputs since there are still notable misclassifications in the last growth stage in September.

Chapter 4 advanced the interpretative analysis conducted in Chapter 3 by calculating the average gradients from pixel weights in input image patches. This calculation is performed through back-propagation in the 3D-ConvSTAR model to create saliency maps for each crop class. These maps revealed spatial locations of significance for each crop type within the testing samples (Figure 4-10). Particularly for maize and soybean, the configurations of the most important pixel clusters suggest intercropping practices, as evidenced by the complementary patterns of important pixels of the two crop types. Conversely, the distinct separation of the pronounced pixels for wheat indicates wheat croplands are comparatively isolated. For other crops, the scattered and less defined pattern of pixel importance corresponds to the model's reduced accuracy and confidence in classifying mixed crop types. Additionally, the soft outputs from the 3D-ConvSTAR model's final Softmax layer, particularly the prediction scores represented in Figure 4-9, indicate a higher confidence level in mapping maize, soybean, and wheat. However, this confidence is relatively low for other crops. Despite having a larger training sample size than wheat, other crops exhibit a mean prediction score of only 71%. This discrepancy highlights a challenge in the model's ability to generalise across a more diverse set of crop types.

In Chapter 5, the interpretation of temporal feature importance over time is achieved by calculating class-specific gradients with respect to time series inputs from the Transformer-AtLSTM (Ensemble) model (Figures 5-12 – 5-13). This analysis examines how the proposed model adapts to interannual and spatial variability. The AtLSTM component within the model prioritised certain periods, especially during late growth and harvesting stages, to pinpoint key phenological stages essential for differentiating crop types. This preference is reflected in the smooth distribution curves of feature importance across crop types, demonstrating the sequential and accumulative nature of the AtLSTM model in capturing temporal dependencies. This finding is consistent with the insights gathered from interpreting the attention LSTM in Chapter 3. Conversely, the Transformer model, with its self-attention mechanism, operates

sequential data processing and identifies temporal dependencies among all pairs of time positions. The resulting feature importance distribution, although less smooth than AtLSTM's, leads to a complementary effect on temporal feature learning. The synergy of these two models is particularly evident in the gradient distributions for InSAR coherence and spectral information of crop classes, which also contributes to the model's transferability to spatial and temporal variations of crops across different geographical contexts. Additional analysis of the importance of input features, particularly focusing on the combined use of m-chi decomposition features and multispectral bands, is further detailed in Figures S11 and S12.

In summary, this research explored the interpretability of various deep learning models applied to crop mapping. It focused on understanding feature learning behaviours, analysing prediction confidence, examining saliency maps, and calculating the importance of input features. It ensures that the developed frameworks and architectures are not only effective in terms of accuracy but are also interpretable through a multi-view understanding of how different features and temporal dynamics are influential to crop identification, which demonstrates the most informative periods, input features, and the complementary nature of ensemble modelling for enhanced crop classification.

6.4 Research Limitations and Recommendations

This research developed deep learning methods for crop mapping using SAR and optical imagery. However, several aspects of the proposed approaches need to be further investigated. A primary limitation of this research is its localized focus on Bei'an county across the three experimental chapters, which does not adequately address the spatial transferability of the models. Within this thesis, Chapter 5 attempted to test the spatiotemporal transferability of the models, specifically in local regions characterized by unique intercropping patterns, but both training and testing samples were collected from the same overall study area. The pre-trained models for crop mapping were tailored to the unique agricultural practices, crop types, and seasonal variations observed within this geographical boundary (Bei'an). Therefore, its current configuration is optimized to generate reliable outputs within this defined area. This limits the scope of the model's applicability and requires the need for additional validation of the pre-trained model in geographically diverse and unseen regions beyond where the training samples were originally collected. Additionally, the models were trained to identify and map a few crops, including maize, soybeans, and wheat. These selections were made based on the predominant

agricultural outputs of the training region. The models' accuracy in identifying or distinguishing other crop types not included in the training data may be reduced. To extend the model's utility beyond its current geographical and crop-type limitations, future research should aim to test model transferability across different geographical areas by training the model with denser remote sensing data and incorporating a broader range of crop classes, which would enable the model to adapt to a wide array of environmental conditions and crop types. This is essential to ensure the robustness and generalisability of the developed methods, particularly for accurately identifying croplands with complex cropping patterns in real-world scenarios.

Future research should consider early-season crop mapping, which seeks to determine the earliest identifiable time for crop classification before harvest. In this thesis, the crop maps were generated based on the full-season image time series in each experimental chapter. This post-harvest classification strategy does not account for the earliest recognisable timings within the growth season. Given the increasing demand in current agricultural production for timely crop mapping, it is important to identify crop types in their early growing season, instead of post-harvest. For example, You and Dong (2020) demonstrated the feasibility of early prediction by using Sentinel-1/2 data to identify rice during its late transplanting period, 120 days before harvest, and to detect maize when it begins heading stage, 60 days before harvest. Timely crop mapping of this nature would greatly benefit the optimisation of cropland planning, as it provides near-real-time information on crop acreage and spatial distribution so that policymakers could deliver informed agricultural decision-making.

The deep learning architectures examined in this thesis, while effective, have inherent limitations. The ensemble learning approaches in Chapters 3 and 5, and the hybrid learning strategy in Chapter 4 involve training multiple models simultaneously, which are computationally intensive. Although these models effectively leverage the strengths of each constituent model, they introduce complexity and require careful consideration of how these models are interrelated to reduce computational power and memory. The Transformer-AtLSTM (Ensemble), for example, employs attention mechanisms in parallel computing. However, it may encounter challenges with the quadratic increase in attention scores from the attention layers processing long input sequences. Generally, deep learning models are data-driven and often subject to resource demands.

Future research should explore integrating theory-driven models, such as physical models, with deep learning frameworks, potentially yielding more accurate and reliable crop map products. Physics-based models are built on established theories and equations that can interpret the physical, chemical, or biological processes involved. However, these models often encounter challenges due to uncertain parameterisations and insufficient representations of the land cover processes. They also require initial conditions to deploy simulation, considering the starting state and boundary conditions that define spatial limits, typically based on observations, experiments, or previous research. In contrast, deep learning models excel at handling nonlinear mapping problems but lack physical insights and suffer from issues such as low interpretability, failure when applied outside of sample conditions, and high demand for a substantial number of training data. To leverage the strengths of both physical and deep learning models, studies have developed hybrid methods that incorporate physical information into deep learning models (Liu et al., 2022; Slater et al., 2023; Li et al., 2024).

This hybrid modelling framework in crop mapping could utilise the outputs of physical models as direct inputs into deep learning models. For example, it could incorporate detailed daily growth measurements under changing environmental conditions, which are produced by physical models based on vegetation scattering properties. These deep learning models can then utilise these inputs along with remote sensing data to make real-time predictions and produce highly accurate crop-type maps that are dynamically responsive to spatiotemporal variations. This hybrid combination could also potentially involve using physical models to simulate conditions absent in the training data or to update predictions based on adding new environmental data, thereby refining predictions of crop types across different regions. Thus, physical models, with their transparent structures and representations of physical processes, complement deep learning models for enhancing crop classification performance, achieving a promising way of synergy and knowledge-guided deep learning in geosciences (Camps-Valls et al., 2021). In summary, adapting the model framework and selecting optimal input features are task-specific, and remain an iterative and challenging process.

References

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

Camps-Valls, G., Tuia, D., Zhu, X.X. and Reichstein, M. eds., 2021. Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science and geosciences. *John Wiley & Sons*.

Dingle Robertson, L., McNairn, H., Jiao, X., McNairn, C. and Ihuoma, S.O., 2022. Monitoring crops using compact polarimetry and the RADARSAT constellation mission. *Canadian Journal of Remote Sensing*, *48*(6), pp.793-813.

Dou, P., Shen, H., Li, Z. and Guan, X., 2021. Time series remote sensing image classification framework using combination of deep learning and multiple classifiers system. *International Journal of Applied Earth Observation and Geoinformation*, 103, p.102477.

Hamad, R.A., Yang, L., Woo, W.L. and Wei, B., 2020. Joint learning of temporal models to handle imbalanced data for human activity recognition. *Applied Sciences*, *10*(15), p.5293.

Hao, P., Di, L., Zhang, C. and Guo, L., 2020. Transfer Learning for Crop classification with Cropland Data Layer data (CDL) as training samples. *Science of The Total Environment*, *733*, p.138869.

He, C., He, B., Tu, M., Wang, Y., Qu, T., Wang, D. and Liao, M., 2020. Fully convolutional networks and a manifold graph embedding-based algorithm for polsar image classification. *Remote Sensing*, *12*(9), p. 1467.

Hu, Q., Sulla-Menashe, D., Xu, B., Yin, H., Tang, H., Yang, P. and Wu, W., 2019. A phenology-based spectral and temporal feature selection method for crop mapping from satellite time series. *International Journal of Applied Earth Observation and Geoinformation, 80*, pp.218-229.

Interdonato, R., Ienco, D., Gaetano, R. and Ose, K., 2019. DuPLO: A DUal view Point deep Learning architecture for time series classificatiOn. *ISPRS Journal of Photogrammetry and Remote Sensing*, *149*, pp.91-104.

Ji, S., Zhang, C., Xu, A., Shi, Y. and Duan, Y., 2018. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sensing*, *10*(1), p.75.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *Nature*, 521(7553), pp.436-444.

Liao, C., Wang, J., Xie, Q., Baz, A.A., Huang, X., Shang, J. and He, Y., 2020. Synergistic use of multi-temporal RADARSAT-2 and VENμS data for crop classification based on 1D convolutional neural network. *Remote Sensing*, 12(5), p.832.

Li, L., Dai, Y., Wei, Z., Shangguan, W., Wei, N., Zhang, Y., Li, Q. and Li, X.X., 2024. Enhancing Deep Learning Soil Moisture Forecasting Models by Integrating Physics-based Models. *Advances in Atmospheric Sciences*, pp.1-16.

Liu, L., Xu, S., Tang, J., Guan, K., Griffis, T.J., Erickson, M.D., Frie, A.L., Jia, X., Kim, T., Miller, L.T. and Peng, B., 2022. KGML-ag: a modeling framework of knowledge-guided machine learning to simulate agroecosystems: a case study of estimating N 2 O emission using data from mesocosm experiments. *Geoscientific Model Development*, *15*(7), pp.2839-2858.

Mahdianpari, M., Mohammadimanesh, F., McNairn, H., Davidson, A., Rezaee, M., Salehi, B. and Homayouni, S., 2019. Mid-season crop classification using dual-, compact-, and full-polarization in preparation for the Radarsat Constellation Mission (RCM). *Remote Sensing*, *11*(13), p.1582.

Nowakowski, A., Mrziglod, J., Spiller, D., Bonifacio, R., Ferrari, I., Mathieu, P.P., Garcia-Herranz, M. and Kim, D.H., (2021). Crop type mapping by using transfer learning. *International Journal of Applied Earth Observation and Geoinformation*, *98*, p.102313.

Pelletier, C., Webb, G.I. and Petitjean, F., 2019. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, *11*(5), p.523.

Raney, R.K., 2019. Hybrid dual-polarization synthetic aperture radar. *Remote Sensing*, *11*(13), p.1521.

Slater, L. J., and Coauthors, 2023: Hybrid forecasting: Blending climate predictions with AI models. *Hydrology and Earth System Sciences*, *27*, pp.1865−1889.

Roy, S.K., Krishna, G., Dubey, S.R. and Chaudhuri, B.B., 2019. HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, *17*(2), pp.277-281.

Rußwurm, M. and Körner, M., 2020. Self-attention for raw optical satellite time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *169*, pp. 421-435.

Sonobe, R., 2019. Parcel-based crop classification using multi-temporal TerraSAR-X dual polarimetric data. *Remote Sensing*, *11*(10), p. 1148.

Tao, C., Meng, Y., Li, J., Yang, B., Hu, F., Li, Y., Cui, C. and Zhang, W., 2022. MSNet: multispectral semantic segmentation network for remote sensing images. *GIScience & Remote Sensing*, *59*(1), pp.1177-1198.

Teimouri, M., Mokhtarzade, M., Baghdadi, N. and Heipke, C., 2022. Fusion of time-series optical and SAR images using 3D convolutional neural networks for crop classification. *Geocarto International*, *37*(27), pp.15143-15160.

Turkoglu, M.O., D'Aronco, S., Perich, G., Liebisch, F., Streit, C., Schindler, K. and Wegner, J.D., 2021. Crop mapping from image time series: Deep learning with multi-scale label hierarchies. *Remote Sensing of Environment*, *264*, p.112603.

Van der Maaten, L. and Hinton, G., 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(11).

Xie, Q., Wang, J., Liao, C., Shang, J., Lopez-Sanchez, J.M., Fu, H. and Liu, X., 2019. On the use of Neumann decomposition for crop classification using multi-temporal RADARSAT-2 polarimetric SAR data. *Remote Sensing*, *11*(7), p.776.

Xu, J., Yang, J., Xiong, X., Li, H., Huang, J., Ting, K.C., Ying, Y. and Lin, T., 2021. Towards interpreting multi-temporal deep learning models in crop mapping. *Remote Sensing of Environment*, *264*, p.112599.

Yang, S., Gu, L., Li, X., Jiang, T. and Ren, R., 2020. Crop classification method based on optimal feature selection and hybrid CNN-RF networks for multi-temporal remote sensing imagery. *Remote Sensing*, *12*(19), p.3119.

You, N. and Dong, J., 2020. Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, *161*, pp.109-123.

You, N., Dong, J., Huang, J., Du, G., Zhang, G., He, Y., Yang, T., Di, Y. and Xiao, X., 2021. The 10-m crop type maps in Northeast China during 2017–2019. *Scientific Data*, *8*(1), p.41.

You, N., Dong, J., Li, J., Huang, J. and Jin, Z., 2023. Rapid early-season maize mapping without crop labels. *Remote Sensing of Environment*, *290*, p.113496.

Zhang, C., Di, L., Hao, P., Yang, Z., Lin, L., Zhao, H. and Guo, L., 2021. Rapid in-season mapping of corn and soybeans using machine-learned trusted pixels from Cropland Data Layer. *International Journal of Applied Earth Observation and Geoinformation*, *102*, p.102374.

Zhang, C., Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J. and Atkinson, P.M., 2018. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *140*, pp.133-144.

Zhong, L., Hu, L. and Zhou, H., 2019. Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, *221*, pp.430-443.

Chapter 7 Conclusions

This thesis presents innovative deep learning methods integrated with multi-source satellite data, aimed at enhancing crop mapping across various scales in Bei'an, China. As outlined in Section 1.5, each research objective addresses progressively intricate scenarios, which range from large-scale crop mapping to dealing with misclassifications caused by local cropping patterns. The objectives are achieved as follows:

Objective 1 - to develop a framework for multi-temporal crop mapping in Bei'an County by using polarimetric SAR-derived data combined with deep learning methods:

Chapter 3 introduced a synergistic approach, the joint ensemble learning of two temporal models – Conv1D and attention-based LSTM – for extracting multi-temporal features from Sentinel-1 dual-pol polarimetric SAR data. This model architecture exhibited superior performance in predicting county-level crop classes in Bei'an for the year 2017. This model, with its optimal selection of SAR features, especially m-chi decomposition, demonstrated its feasibility in managing imbalanced class distribution and efficiently distinguished between major and minor crops with similar phonologies. This chapter strengthened the importance of using long time series data in crop classification (as illustrated in Table 3-3 and Figure 3-17 for in-season crop mapping) and revealed the model's ability to capture temporal dependencies in multi-temporal SAR data (Figures 3-14 to 3-16).

Objective 2 - to construct a sophisticated deep learning architecture that combines multiple models for county-level crop mapping based on the fusion of multi-temporal optical and SAR datasets for Bei'an County:

In Chapter 4, a novel framework was developed that integrates Sentinel-1 polarimetric features with Sentinel-2 multispectral reflectance. This novel hybrid model architecture, namely 3D-ConvSTAR, connects 3D-CNN layers with convolutional recurrent layers (ConvSTAR). It demonstrated enhanced performance in crop mapping, particularly in its effectiveness in extracting features from the combined SAR and optical datasets, presenting a clear advantage over using SAR data alone. This architecture also showcased its robustness against imbalanced class distributions and outperformed other data augmentation techniques in performance, as detailed in Table 4-3. Although it faced challenges with increased training parameters and

limitations in classifying underrepresented crops, the model still exhibited potential in crop classification, notably in terms of prediction confidence and saliency maps, as illustrated in Figures 4-9 and 4-10.

Objective 3 - to design a deep learning based approach tailored for mapping areas of intercropping in Bei'an using interferometric SAR coherence and high resolution (5m) multispectral data:

Chapter 5 proposed an innovative multi-branch deep learning architecture that integrates InSAR coherence with multispectral bands. This approach, designed to leverage the synergy of multi-source data, deep learning, and machine learning models, assembles in the Transformer-AtLSTM-RF model. Enhanced by a rule-based decision fusion technique, this model effectively maps crops in smallholder croplands characterized by complex intercropping systems. It achieved higher transfer learning accuracy across various time frames and geographical locations within Bei'an, compared to standalone temporal and FCN-based model architectures (Tables 5-2 to 5-5). The chapter also highlighted the importance of fine-tuning pre-trained classifiers with representative data during transfer learning, evaluating the input feature importance through models' decision-making process (Figures 5-12 to 5-13), and the benefits of using multi-source data such as the fusion of InSAR coherence and multispectral data.

Future research directions aim to optimise these deep learning architectures for increased efficiency, generalisability, and transferability. This includes exploring higher spatial resolution (> 5m) data, denser remote sensing data, and incorporating environmental variables such as temperature, precipitation, or soil moisture as predictors. Further, there is a need to collect data samples that are more representative of specific agricultural conditions, especially for areas with complex cropping patterns. The goal is to develop more robust models capable of handling diverse agricultural scenarios, ensuring accuracy and efficiency in producing crop map products across different regions at various scales.

Supplementary Material

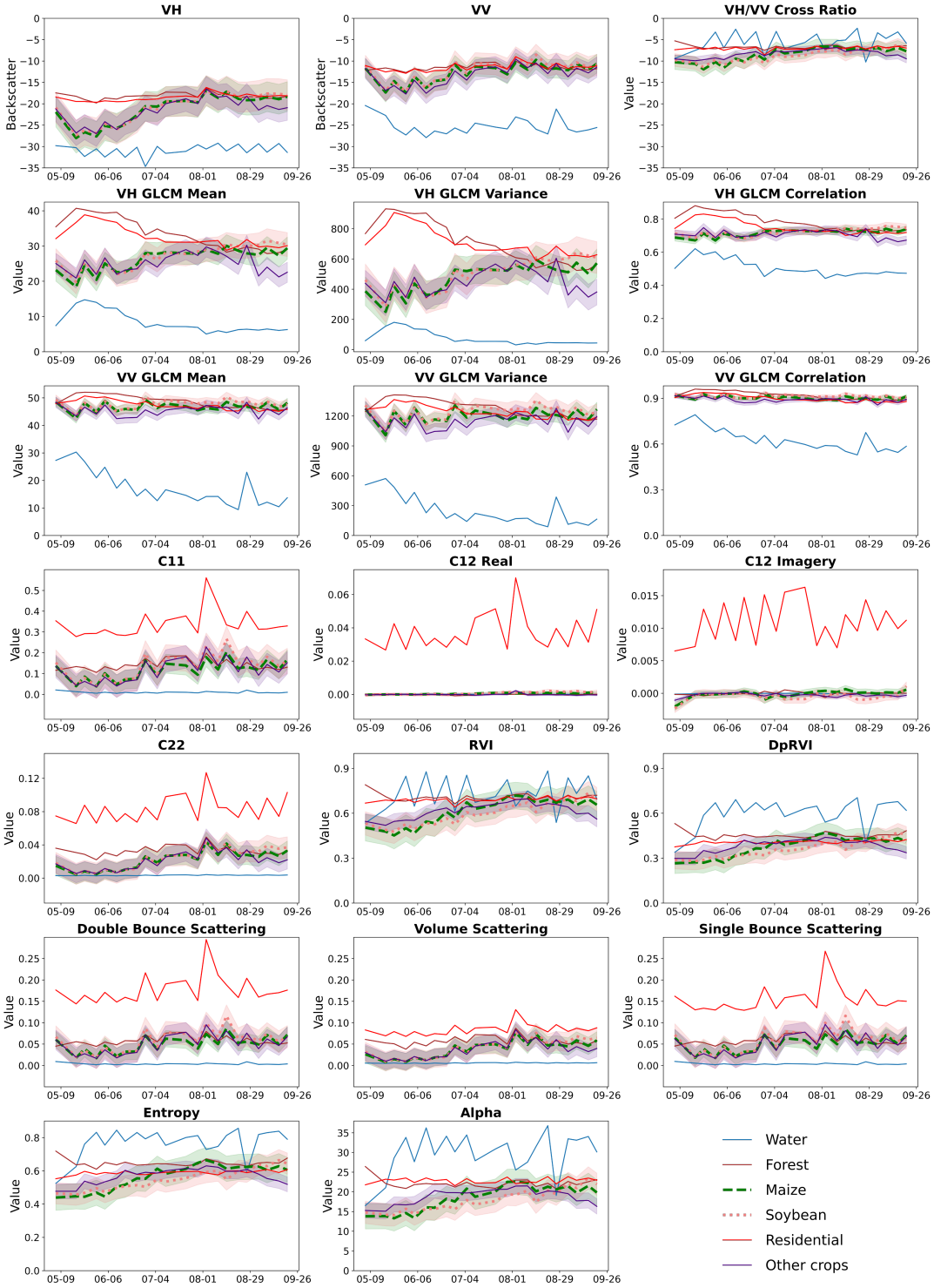

Figure S1. Temporal profiles of Sentinel-1 backscatter (VV and VH), cross-ratio (VH/VV), covariance matrix parameters (C11, C12 Real, C12 Imagery and C22), SAR vegetation indices (RVI and DpRVI), GLCM and polarimetric decomposition features. The buffer area in each subplot is one standard deviation from the mean for each crop type.

Figure S2. Feature selection based on Boruta. Circles indicate the features with the highest importance scores, while full-time features are marked with stars. GLCM_c, GLCM_v and GLCM_m indicate correlation, variance and mean.

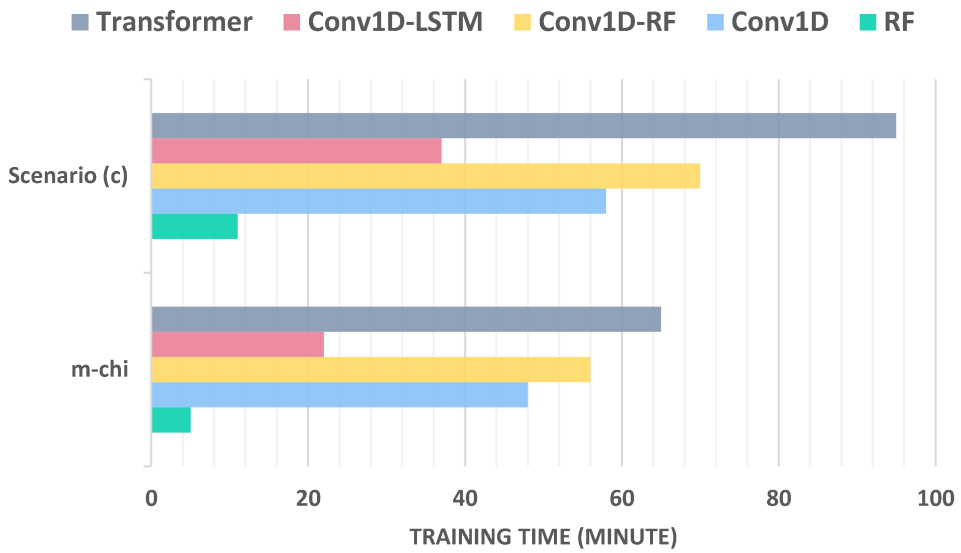

Figure S3. The comparison of model training time for m-chi features and scenario (c) based on 10% of ground-truth training samples.

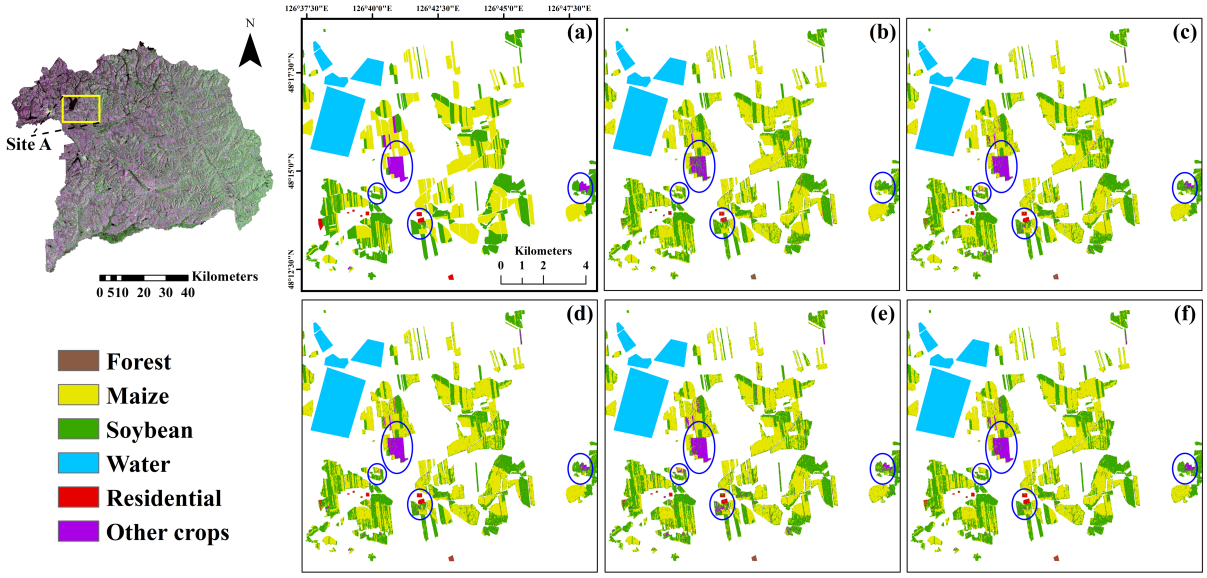

Figure S4. Comparison of map predictions for Site A. (a) Reference labels, (b) RF, (c) Conv1D, (d) Conv1D-RF, (e) Transformer, (f) Conv1D-LSTM. Sentinel-1 false colour composite (Blue: single bounce scattering, Green: double bounce scattering, Red: volume scattering).

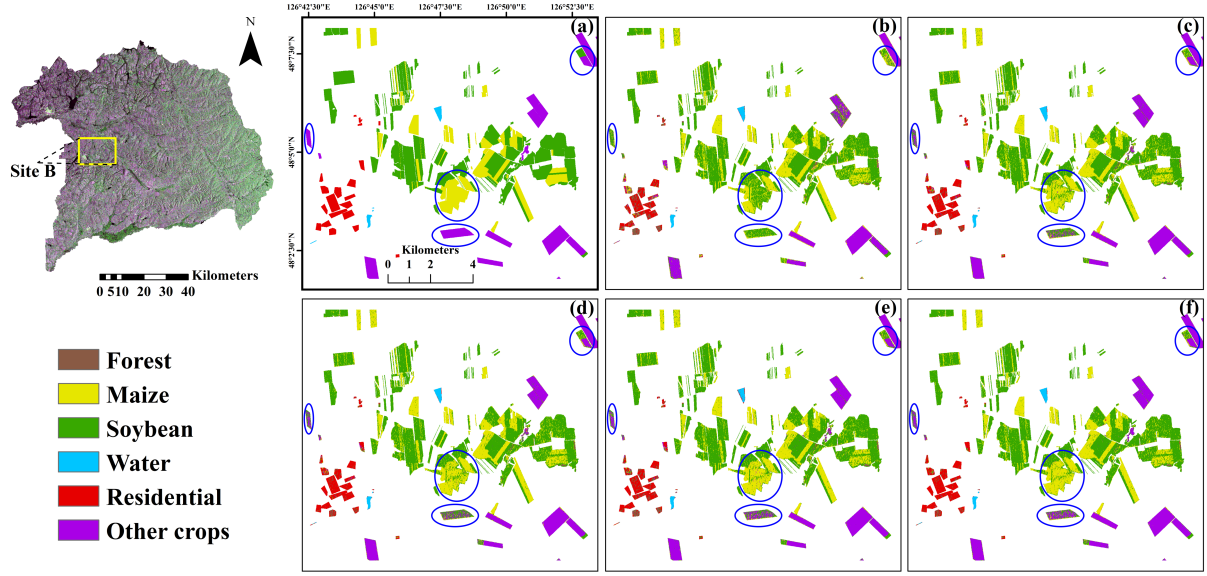

Figure S5. Comparison of map predictions for Site B. (a) Reference labels, (b) RF, (c) Conv1D, (d) Conv1D-RF, (e) Transformer, (f) Conv1D-LSTM. Sentinel-1 false colour composite (Blue: single bounce scattering, Green: double bounce scattering, Red: volume scattering).

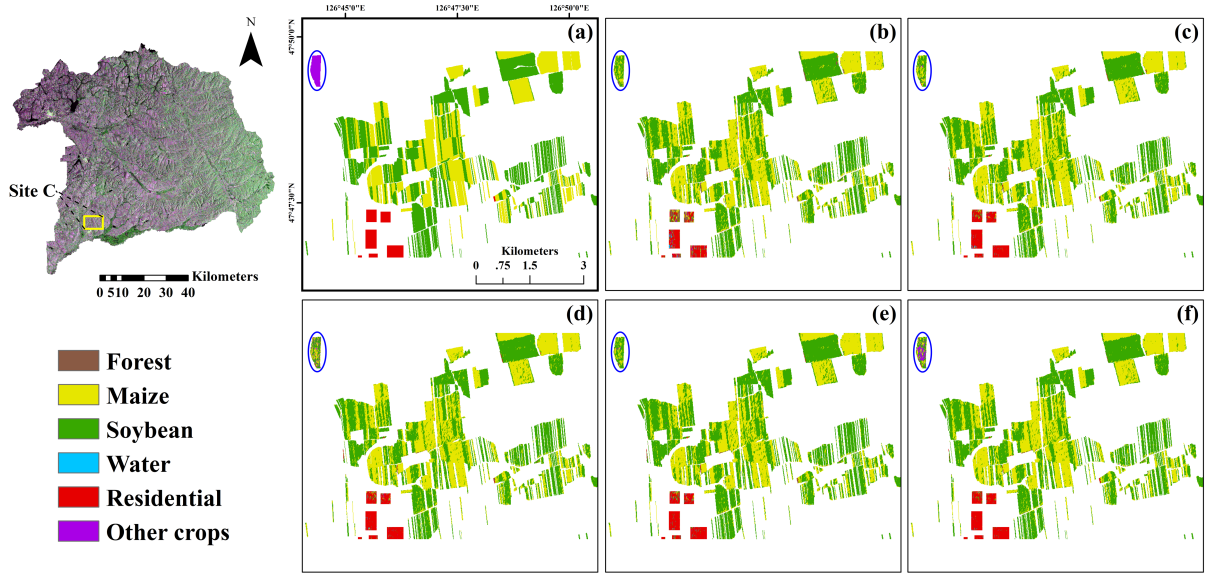

Figure S6. Comparison of map predictions for Site C. (a) Reference labels, (b) RF, (c) Conv1D, (d) Conv1D-RF, (e) Transformer, (f) Conv1D-LSTM. Sentinel-1 false colour composite (Blue: single bounce scattering, Green: double bounce scattering, Red: volume scattering).

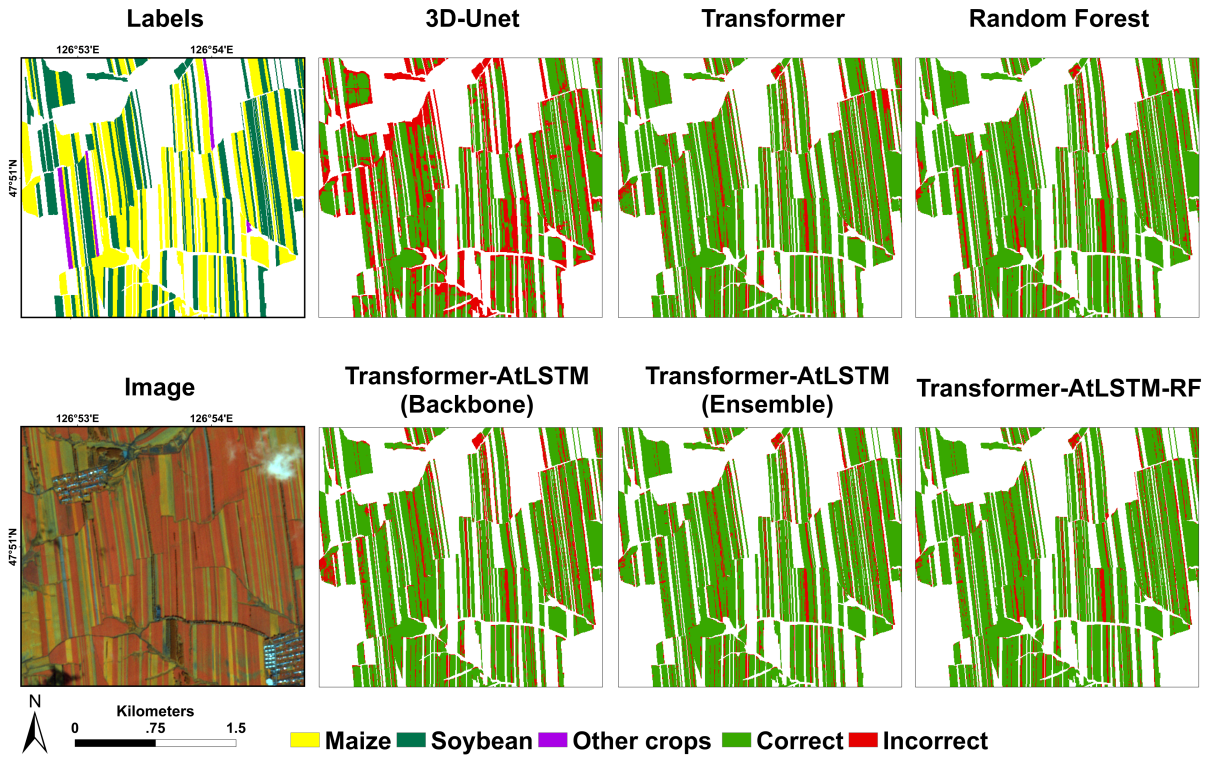

Figure S7. Enlarged extent of crop mapping results for Site A 2018. Captions follow Figure 5-9.

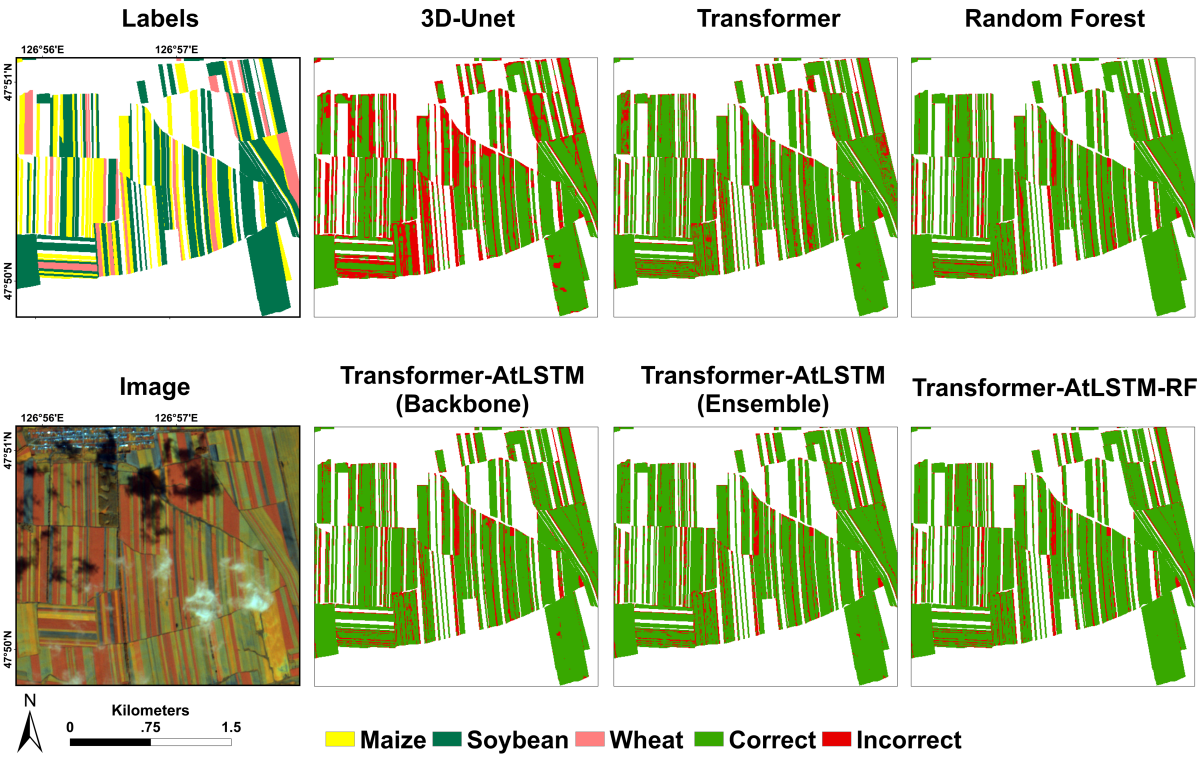

Figure S8. Enlarged extent of crop mapping results for Site B 2018. Captions follow Figure 5-9.

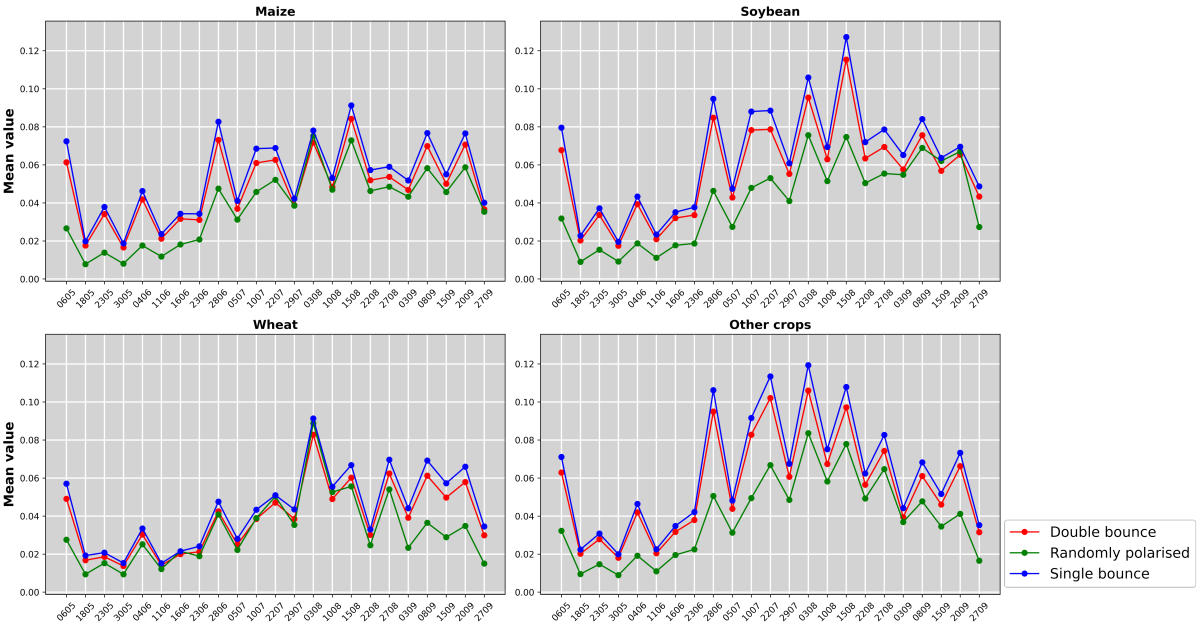

Figure S9. Temporal profiles of crops based on Sentinel-1 m-chi decomposition features in Bei'an 2017.

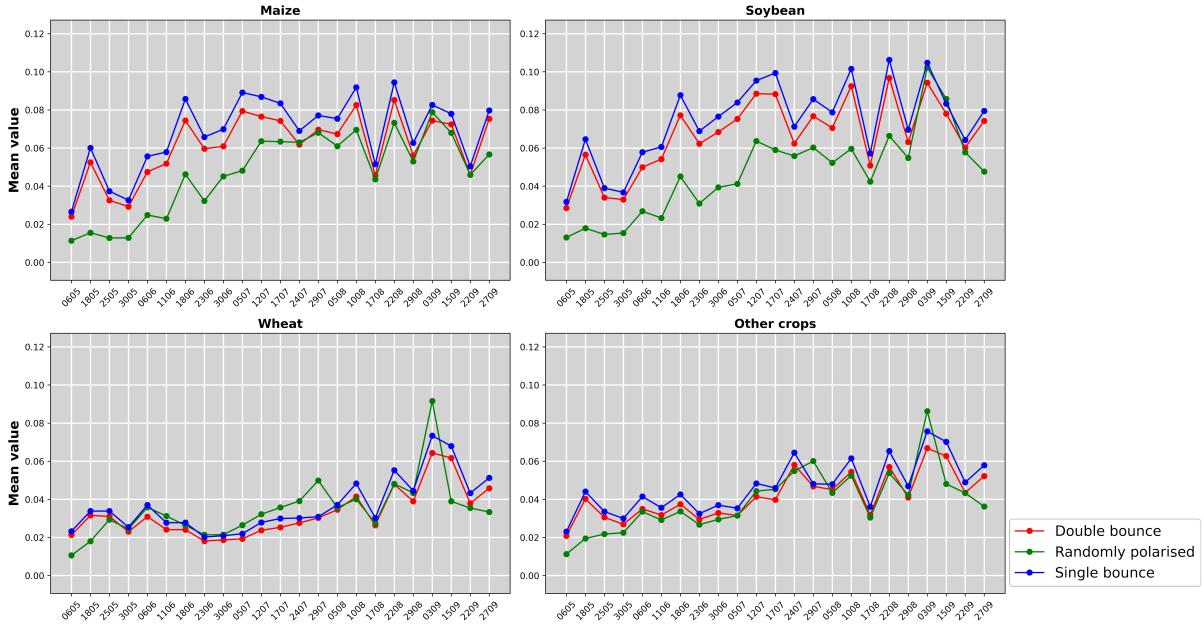

Figure S10. Temporal profiles of crops based on Sentinel-1 m-chi decomposition features in Bei'an 2018.

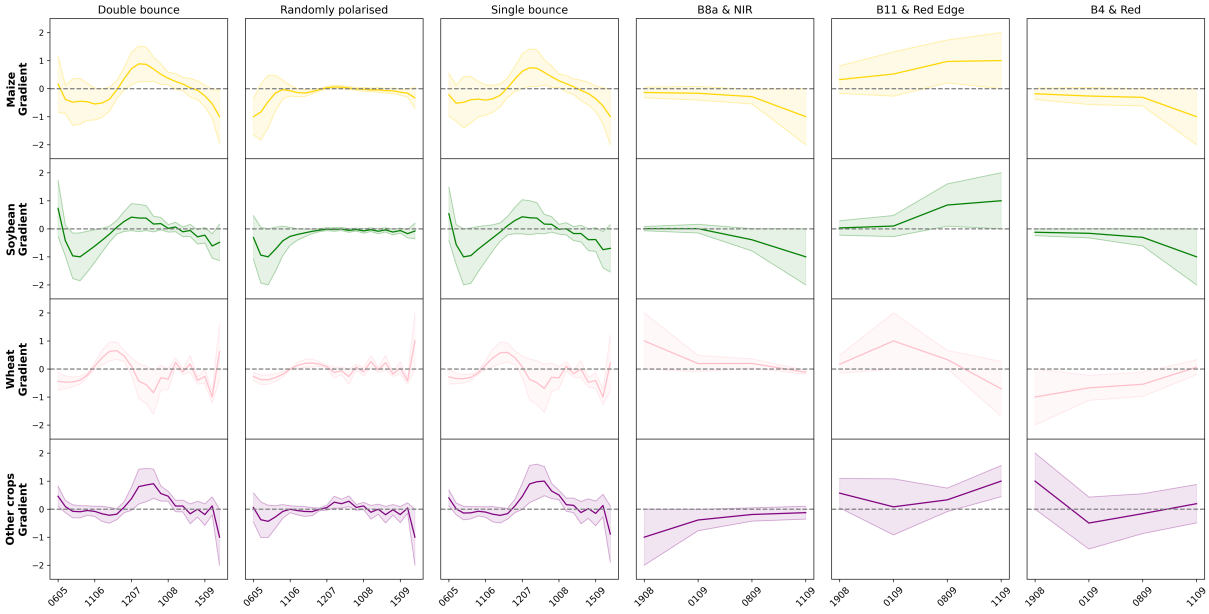

Figure S11. Average gradients of attention weights with respect to inputs from the AtLSTM end. 3000 samples of Site A and B were randomly selected from the attention weight layer for visualisation. The light buffer areas indicate one standard deviation from the average value. Positive (negative) values indicate a positive (negative) correlation between the predicted score and input features and vice versa. The value range was scaled from -1 to 1. Predictors are m-chi decomposition features.

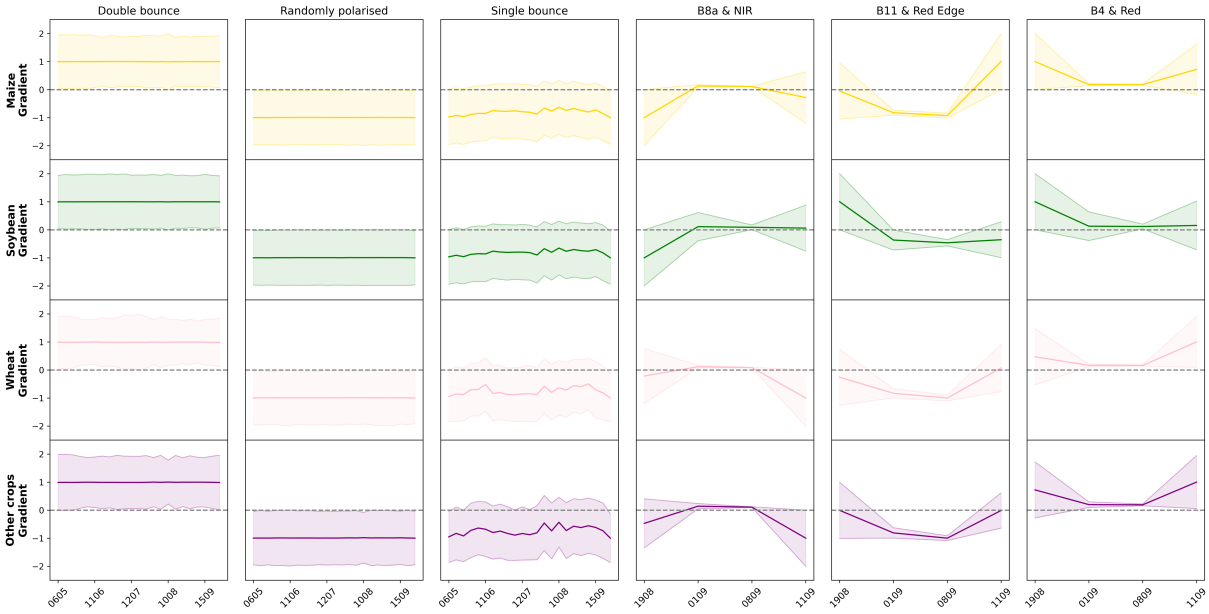

Figure S12. Average gradients of attention weights with respect to inputs from the Transformer end. 3000 samples of Site A and B were randomly selected from the second self-attention layer for visualisation. The light buffer areas indicate one standard deviation from the average value. Positive (negative) values indicate a positive (negative) correlation between the predicted score and input features and vice versa. The value range was scaled from -1 to 1. Predictors are m-chi decomposition features.

Table S1. F1 score per class, mean F1 score and overall accuracy (OA) from the 2017 testing set. The best values are highlighted in bold in the columns.

Model	Maize (%)	Soybean (%)	Wheat (%)	Other crops (%)	Mean F1 (%)	OA (%)
3D U-Net	91.0	**93.1**	81.1	**68.5**	**83.4**	**88.0**
Transformer	91.0	80.6	67.2	60.4	74.8	82.2
AtLSTM	91.4	84.6	85.6	64.8	81.6	85.7
RF	**92.2**	86.0	**86.7**	64.8	82.4	87.1

Table S2. IoU per class and mean IoU (mIoU) from the 2017 testing set. The best values are highlighted in bold in the columns.

Model	Maize (%)	Soybean (%)	Wheat (%)	Other crops (%)	mIoU (%)
3D U-Net	83.5	**87.1**	68.1	**52.1**	**72.7**
Transformer	83.5	67.6	50.6	43.3	61.2
AtLSTM	84.1	73.3	74.8	47.9	70.0
RF	**85.5**	75.5	**76.5**	47.9	71.3

Table S3. Transfer Site A: overall accuracy (OA) and mean F1 score for 2018 crop mapping validation. The best values are highlighted in bold in the columns.

Model	Input feature	Maize (%)	Soybean (%)	Other crops (%)	Mean F1 (%)	OA (%)
Transformer-AtLSTM (Ensemble)	M-chi + Optical	95.4	95.9	63.3	84.9	94.9
	Coherence + Optical	**96.1**	**96.7**	**77.4**	**90.1**	**96.1**

Table S4. Transfer Site A: IoU and mean IoU (mIoU) of 2018 crops. The best values are highlighted in bold in the columns.

Model	Input feature	Maize (%)	Soybean (%)	Other crops (%)	mIoU (%)
Transformer-AtLSTM (Ensemble)	M-chi + Optical	91.2	92.2	46.4	76.6
	Coherence + Optical	**92.6**	**93.6**	**63.1**	**83.1**

Table S5. Transfer Site B: overall accuracy (OA) and mean F1 score for 2018 crop mapping validation. The best values are highlighted in bold in the columns.

Model	Input feature	Maize (%)	Soybean (%)	Wheat (%)	Other crops (%)	Mean F1 (%)	OA (%)
Transformer-AtLSTM (Ensemble)	m-chi + Optical	86.1	91.1	73.8	78.4	82.4	87.5
	Coherence + Optical	**88.5**	**92.6**	**78.7**	**84.9**	**86.2**	**89.9**

Table S6. Transfer Site B: IoU and mean IoU (mIoU) of 2018 crops. The best values are highlighted in bold in the columns.

Model	Input feature	Maize (%)	Soybean (%)	Wheat (%)	Other crops (%)	mIoU (%)
Transformer-AtLSTM (Ensemble)	m-chi + Optical	75.6	83.7	58.4	64.5	70.6
	Coherence + Optical	**79.4**	**86.2**	**64.8**	**73.7**	**76.0**