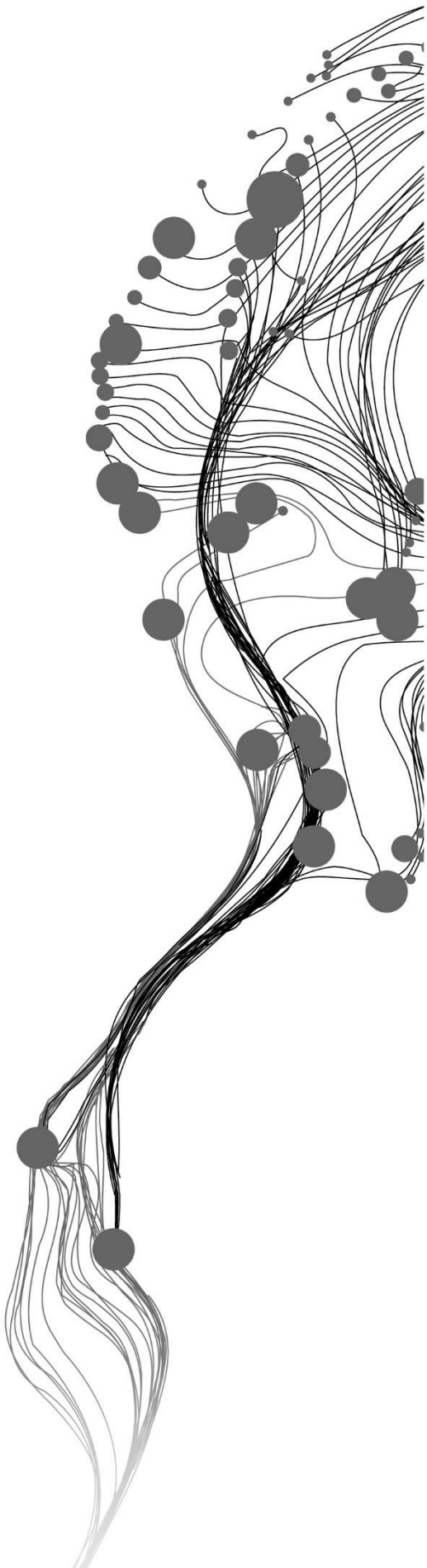


CROP TYPE IDENTIFICATION AND YIELD ESTIMATION USING MULTI-TASK LEARNING CNN AND SWIN ARCHITECTURES

MOHAMED.M.A. GAMIL

June 2024

SUPERVISORS:
Dr. Mahdi Farnaghi
Prof. Dr. Raul Zurita Milla

CROP TYPE IDENTIFICATION AND YIELD ESTIMATION USING MULTI-TASK LEARNING CNN AND SWIN ARCHITECTURES

MOHAMED.M.A. GAMIL
Enschede, The Netherlands, June 2024

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

SUPERVISORS:
Dr. Mahdi Farnaghi
Prof. Dr. Raul Zurita Milla

THESIS ASSESSMENT BOARD:
Chair: Dr. F.O. Ostermann
External examiner: Dr. R. Vargas Maretto
Supervisors: Dr. Mahdi Farnaghi & Prof. Dr. Raul Zurita Milla

ABSTRACT

Early crop yield estimation (i.e., before the harvest season) is essential for effective commodity market management, ensuring food security and understanding various crop production trends at preliminary stages. Although several crops can be cultivated in a given area, most research focuses on single-crop yield estimation using Earth Observation (EO) data and Artificial Intelligence (AI) models. However, developing these crop-specific models is time-consuming and computationally expensive. A classified image of crop types helps to improve the accuracy of multi-crop yield estimation models because it guides the model to focus on the relevant images' regions. However, such a layer is not always available because its creation is expensive and time-consuming. Furthermore, the creation of such a crop-type layer requires a substantial amount of labelled ground truth data that is often not available. To estimate multi-crop yields while addressing the unavailability of a classified crop type layer, we developed multi-task learning models for crop type identification and yield estimation. We hypothesized that multi-task learning models can learn both tasks concurrently. Additionally, these models solely require satellite imagery inputs. In this thesis, we estimated the yield of two crops: corn and soybean and the case study covered the top four states in the USA in corn and soybean production namely Indiana, Iowa, Illinois, and Minnesota. Consequently, we developed two base CNN models for multi-crop yield estimation. The first one only used sentinel-2 images with eight bands and the second one used a classified crop layer (known as CDL) as an additional band to demonstrate the role of the CDL in achieving higher accuracies. Regarding the multi-tasks approach, we developed two models: one based on U-net and the other utilizing the Swin transformer-based architecture. The two multi-task learning models showed promising results in multi-crop yield estimation. For instance, both models achieved yield estimation accuracies comparable to the CNN model that relies on CDL as input. Additionally, those models proved their applicability for multi-crop yield estimation while solving the lack of a CDL-like layer in many other countries. However, there is still room for further improvements to increase the models' accuracies. These improvements relate to date, data acquisition and model architecture aspects. For instance, for date-related enhancements, incorporating temporal data in different years and different times in the mid-season of the crops could be beneficial. Regarding data acquisition, integrating different bands of sentinel-2, different sensors' data, and additional parameters such as weather data, could improve the models' generalizability. For model-related improvements, different loss functions and optimization techniques could improve performance. Finally, the primary contribution of this research lies in the development of multi-task learning models for crop type identification and yield estimation. These models proved to be actionable models that can be used directly to estimate multi-crop yields without the need for CDL while still achieving good results. This approach addressed the limitations in developing crop-specific models by reducing the number of parameters to be learned and the computation time and resources required.

ACKNOWLEDGEMENTS

I would like to express my profound gratitude to my first supervisor, Dr. Mahdi Farnaghi. Dr. Mahdi, you have been more than just a supervisor; you have been a mentor whose guidance and support have been invaluable throughout my MSc journey. I truly appreciate your patience and the generous time you have given me. Your open-minded approach and encouragement to think creatively have inspired me to employ state-of-the-art methods, expanding the boundaries of my research. Your confidence in my abilities and your insightful advice have been fundamental to my academic and personal growth.

I would also like to extend my sincere thanks to my second supervisor, Prof. Dr. Raul Zurita Milla. Prof. Raul, your critical thoughts, and advice have significantly shaped the quality of my work. Your understanding and appreciation of my efforts have motivated me to strive for excellence. The insightful ideas you provided about the results and your guidance in improving the presentation of my findings have been incredibly valuable. Your constructive feedback has refined my research and made this thesis a robust piece of work.

My gratitude also goes to the chair and the external examiner of my assessment board, Dr. F.O. Ostermann and Dr. R. Vargas Maretto. I am thankful for your time and effort to be part of my thesis assessment.

I am deeply thankful to Dr. Islam Fadel and Dr. Hakan Tanyas for agreeing to use their server to run my models. This thesis would not have been possible without your generous support. Importantly, I extend my heartfelt thanks to Arun Venugopal, whose MSc work from 2023 formed the foundation of my research. Thank you, Arun, for consistently being available to answer my questions and for your unwavering support and assistance whenever needed. Additionally, I would like to express my gratitude to the CRIB team, particularly Dr. Serkan Girgin and Jay Gohil, for providing the platform to run my models and for your help when I faced problems with the installations of some libraries. Your technical support has been instrumental in the successful completion of my research. Moreover, the occasional use of ChatGPT for grammar check and suggesting some word synonyms and speeding up some code writing, was very helpful.

My heartfelt thanks also go to the ITC Excellence Scholarship for their financial support, which made my studies possible. I extend my gratitude to the ITC teachers and personnel for their dedication, guidance, and the high-quality education they provided.

To my friends who stood by me during the challenging period when I stayed in my room for almost two months due to a foot injury, Thanks a million, I will never forget your help. I am especially grateful to Abulraheem Cissey, Ahmed Hemoudi, Mostafa Gomaa, Islam Fadel, Abdullah Banger, Moamen Abayizid, Ramy Rabie and Farag Sayed for their consistent support since I arrived in the Netherlands. Your kindness and companionship, academic and life advice have made this journey much more bearable and joyful.

Lastly, I would like to acknowledge the support of my family and friends who have been a great source of encouragement throughout my MSc journey.

Thank you all for your invaluable contributions and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Background

Crop yield estimation is an important field in precision agriculture. It plays a vital role in estimating future harvests, managing crops to enhance overall productivity, and aligning crop production with market demand (Liakos et al., 2018). It is crucial to estimate the crop yield during the crop season not only for food security purposes but also for commodity market management and to gain insights into the fluctuations in the yield patterns (Desloires et al., 2023). Additionally, enhancing yields through field-level agricultural management is important for tackling worldwide food security concerns (Sagan et al., 2021). This is why, it is paramount to early estimate crop yield and define the variables affecting its fluctuations locally and globally.

Various approaches were sought to estimate crop yield. Historically, estimating crop yield involved extensive field surveys. This process was labour-intensive and required significant time investment (Bi et al., 2023). Then, several methods were explored for estimating crop yield, including statistical models and process-based models. Process-based models encounter challenges because of insufficient parameterization, validation, and calibration data. On the other hand, machine learning models, which fall under statistical models, gained considerable attention due to the progress in big data technologies and high-performance computing (Srivastava et al., 2022). Recently, the availability of EO data in spatial and temporal form, the development of DL models and the advances in computational power have made it easier and less time-consuming. When compared to yield prediction methods based solely on meteorological factors, utilizing remote sensing imagery offers a more comprehensive understanding of the plants' growth status (Bi et al., 2023).

EO data play a crucial role in predicting crop production due to their frequent spatial and temporal image availability (Marshall et al., 2022). Klompenburg et al. (2020) did a systematic literature review about the use of ML models in crop yield prediction and the features used as well. The authors found that the most applied algorithms in descending order were Neural Networks (NNs), Linear Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting. Among the NNs, Deep Neural Networks (DNNs) architectures like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) were widely applied. Moreover, the authors found that the most frequently used features were rainfall, temperature, and soil type. CNNs were mostly incorporated in estimating crop yield with good accuracy in terms of Root Mean Square Error (RMSE). However, lacking enough data for training leads to overfitting and less applicability in practice (Oikonomidis et al., 2023).

In this research, we utilized EO data and deep learning (DL) models to estimate yields for two crops. Corn and soybean were selected as our focus crops. The research encompassed four states in the USA, chosen based on their ranking as the top producers of both corn and soybean in the country, namely Indiana, Iowa, Illinois, and Minnesota.

1.2. Literature Review

1.2.1. EO data, DL, and several factors for single Crop yield estimation

Various ML and DL models have been utilized with EO/UAV data and several environmental and weather factors for single-crop yield estimation. Jhajharia & Mathur. (2023) used several ML models to predict crop yield in Rajasthan, India, using remote sensing and weather data. They used Decision Tree (DT), RF, Support Vector Regression (SVR), and Least Absolute Shrinkage and Selection Operator (LASSO) and proved that RF had the highest accuracy with (R^2: 0.77, RMSE: 0.39 t/ha, MAE: 0.28 t/ha). Zhang et al. (2023) used vegetation indices derived from Landsat 8 and Sentinel-2 images with the Bayesian optimized CatBoost model for robust crop yield estimation, proving superiority over LASSO, SVR, and RF. Lin et al. (2023) utilized sentinel-2 images, CDL, and meteorological data for crop yield estimation at the county level based on a ViT-based model. Htun et al. (2023) utilized sentinel-2 images to create four different indices (NDWI, RGVI, SAVI and NDVI) for Rice yield prediction using a Multiple Regression model. Joshi et al. (2023) integrated Sentinel-1 and Landsat 8 temporal vegetation indices with different ML models (LASSO, SVM and RF) for predicting winter wheat yield. Sun et al. (2019) included various factors to develop and train a DL model for estimating county-level soybean yield during the season and at the end of the season. These factors were MODIS Land Surface Temperature (LST), MODIS Surface Reflectance (SR), weather data, and crop growth variables. Wang et al. (2023) used multi-spatial MODIS satellite images with 3D CNNs to estimate the county-level soybean yield. Johnson et al. (2016) utilized vegetation indices such as NDVI and EVI from MODIS with several linear regression and neural network-based models to estimate the yield of crops such as spring wheat, canola, and barley. Zhang et al. (2021) predicted maize yield in smallholder farms in China using three ML approaches: LSTM, LASSO, and Light Gradient Boosting Machine (LightGBM). Bi et al. (2023) used handheld cameras to acquire time-series images with high resolution and developed a transformer-based model using ViT for soybean yield estimation.

1.2.2. Multi-crop yield estimation

Different crop fields could coexist in the same region and be adjacent in the acquired EO data. However, most of the work in crop yield estimation using DL and EO data focuses on single crops. Developing crop-specific models is computationally inefficient, takes significant time and does not take into consideration the interactivity among different crop types (Khaki et al., 2021). Multi-task learning in crop yield estimation was used to estimate single crop yield while classifying its level or estimating the Grain Protein Content (GPC) (Chang et al., 2024; Z. Sun et al., 2022a). Chang et al. (2024) developed a multi-task learning DL model using UAV images to estimate the rice crop yield. The authors developed a two-head model, one head to estimate the crop yield and the other to classify the level of the yield (high, low). Sun et al. (2022) integrated Lidar data and multi-spectral data in multi-task learning to simultaneously estimate the wheat yield and the GPC. The authors combined losses by adding losses of both tasks with the same weights.

To the best of our knowledge, only one study has focused on estimating both corn and soybean yields simultaneously. Khaki et al. (2021) developed a multi-task learning DL model (YieldNet) with two regression heads, one for corn yield estimation and one for soybean yield estimation. The author designed the backbone of the model (CNN-based) to function as a shared feature extractor for both tasks. Furthermore, the method was evaluated against the same model architecture, first using only the corn head, and then using solely the soybean head. The dual-head model showed higher accuracy for both corn and soybean yield estimation compared to the single-head models. YieldNet was also compared with several single-head models to predict corn and soybean yields separately, and it showed superior accuracy. This improvement is because of the transfer of feature learning between corn and soybean facilitated by the common feature extractor. However, the model still requires CDL as input, which is a limitation since CDL is not available in all countries. This is because collecting CDL through field surveys is time-consuming. Additionally, employing AI models to

generate CDL automatically requires a substantial amount of labelled ground truth data, which is unavailable in many countries (Mohammadi, 2024).
Including CDL as input to the model guides the model to look at the specific crops in the images and ignore the other crops and the background (Venugopal, 2023). Venugopal (2023) experimented with testing a CNN model without adding CDL and with adding CDL for estimating soybean per input image and produced saliency maps for both cases. The CNN model with CDL proved higher accuracy in estimating the crop yield and its saliency maps focused on the soybean fields. On the other hand, the model without CDL looked at different fields and had relatively bad accuracy. Therefore, CDL could also enhance multi-crop yield estimation. Moreover, improving multi-crop yield estimation could be achieved through multi-task learning models to identify the crop type and estimate the crop yield concurrently.

1.2.3. Multi-task learning

Multi-task learning is a way to train models for parallel tasks using common features to enhance all tasks (Caruana, 1997). It uses a common model for different tasks while utilising shared feature representations. In multi-task learning, there are three main pillars: model architecture design, optimization methods, and task-specific learning (Crawshaw, 2020). As explained by Crawshaw. (2020), there are five types of architecture design in computer vision for multi-task learning: Shared Trunk, Cross-Talk, Task Routing, Prediction Distillation, and Single Tasking. The Shared Trunk design is a common and traditional architecture in multi-task learning. It has a global model used for feature extraction that has a single output to be used for multiple tasks. The Cross-Talk design consists of individual networks, one for each task. The information is transferred between the networks through the parallel layers in each task network. Then, the layers' output is linearly combined and fed into the next layers. The Task Routing design offers a less rigid architecture in terms of parameter sharing. Instead of sharing parameters at the layer level, it allows feature-level parameter sharing. The Prediction Distillation design begins with generating initial outcomes for all tasks and then using them to enhance the final predictions. The Single Tasking architecture, unlike the other designs, is developed to perform the inference for one task at a time.

Regarding the optimization methods, Crawshaw (2020) summarized six main categories for multi-task models' optimization techniques: Loss weighting, Regularization, Task scheduling, Gradient modulation, knowledge distillation, and multi-objective optimization. Loss weighting techniques mostly use the weighted average of separate task losses to calculate the overall loss used for the model's backpropagation. Those techniques usually differ in the way of defining the weights. The regularization techniques are usually applied to soft-sharing models that have separate task models and do not share parameters. Task scheduling is used to prioritize which task, or some tasks are being used for the training in each training step. Gradient modulation methods are important in the cases of conflicting gradients of different tasks. knowledge distillation is often used to impart the expertise of multiple single-task networks conventionally named "teacher" into a single multi-task model named "student." Multi-objective optimization is employed to address the weakness of averaging the different tasks' losses into one value because this leads to losing vital information. Therefore, it does not aim to achieve a global minimum, instead, it seeks Pareto optimal solutions. Finally, there are three research paths identified in the third pillar of multi-task learning (Crawshaw, 2020). Firstly, to cluster the tasks into groups that could learn from each other simultaneously. Secondly, to implement techniques to assess when knowledge transfer among tasks enhances learning. The third path focuses on creating an embedding space for the tasks themselves.

Multi-task learning has been used in different fields especially in autonomous driving since it needs real-time actions based on multiple decisions. Ebert et al. (2022) developed a multi-task model for autonomous driving to simultaneously perform semantic segmentation, object detection and human pose estimation using a common backbone and three heads. The combined loss function is a weighted average of the three tasks' losses. This model reduced the learned parameters number and increased the model performance. Furthermore, it reduced the inference time making it convenient for occupancy monitoring. Liu & Wang, (2019) developed AdvNet, which is a multi-task model with a common backbone and two heads, one for lane segmentation and the other for obstacle detection. Cipolla et al. (2017) developed a multi-task learning

model with a common backbone and three different heads, one is for semantic segmentation and the other two are for regression (instance segmentation and depth estimation).

Table 1 summarizes the studies mentioned in the literature review that focus on crop yield estimation with EO data and ML models. It highlights the models used for each study, the number of crops being estimated concurrently, the level to which the crop yield is estimated (county or field level), and the factors used besides the EO data. It is shown that only one study concurrently estimated corn and soybean. Moreover, for studies that used the United States Department of Agriculture (USDA) crop yield data, most of them estimated the crop yield at the county level since the data is available at the county level as BU/acre. However, only one study downsampled the data to the pixel level and estimated single-crop (soybean) total yield to the level of the input images.

Table 1: crop yield estimation studies using EO data and ML models

Literature	Models used	Crops (single or multiple)	Yield estimation level (county or farm)	Factors
Johnson et al. (2016)	Linear, regression, NNs	single crops (spring wheat, canola, and barley)	Census Agricultural Regions (CARs), Canada	NDVI and EVI
Sun et al. (2019)	CNN, LSTM	Single crop(soybean)	county-level, USA	MODIS SR, LST, weather data.
Zhang et al. (2021)	(LASSO), LightGBM, and LSTM	Maize	Field-level	Vegetation indices and weather data
Khaki et al. (2021)	CNNs	Multi-crops (Corn and soybean)	County-level	MODIS data and CDL
Sun et al. (2022)	RNN, LSTM, CNN, attention modules	Single crop (wheat)	Field-level	Lidar and multi-spectral data
Jhajharia & Mathur. (2023)	LASSO regression, SVR, DT, RF	Single crop (wheat)	District-level	NDVI, EVI, LAI, weather data
Htun et al. (2023)	Multiple regression	Single crop (Rice)	Ground reference points (GRPs)	NDWI, RGVI, SAVI and NDVI from sentinel-2 images
Joshi et al. (2023)	(LASSO, SVM and RF	Single crop (Winter wheat)	County-level, USA	Vegetation indices, climatic and soil variables
Bi et al. (2023)	ViT	Soybean	Field level using handheld sensor	Seed information
Zhang et al. (2023)	CatBoost (BO-CatBoost)	Single crop (Winter wheat)	Field-level	Landsat-8 and sentinel-2

	regression model, LASSO, SVM, and RD			vegetation indices
Lin et al. (2023)	ViT	Single crops of (corn, soybean, cotton, winter wheat)	County-level, USA	sentinel-2 images, CDL, and meteorological data
Wang et al. (2023)	3D CNNs	Single crop (soybean)	County-level, USA	multi-spatial MODIS
Chang et al. (2024)	CNN, LSTM	Single crop (Rice)	Field-level	UAV images
Venugopal (2023)	CNN	Single crop (soybean)	Input Image-level	Sentinel-2 images, CDL

1.3. Research Gap

Based on the literature and to the best of the author's knowledge, there are no developed multi-task learning models for multi-crop type identification and yield estimation. Therefore, this research focuses on two important problems. Firstly, most existing studies focus on single-crop yield estimation. However, in practice, different crop fields are often adjacent in the acquired EO data, and developing separate models for each crop type is inefficient and time-consuming, failing to leverage the advantages of transfer learning in developed DL models. Secondly, while various meteorological and environmental factors influence crop yield estimation, CDL is crucial for directing models to relevant image regions. Nonetheless, CDL is not always available across different countries and needs a substantial number of ground truth labels to automatically create CDL using DL models.

1.4. Research Objectives and Questions

The primary objective of this MSc thesis is to estimate multi-crop yields accurately using multi-task learning DL models for simultaneous segmentation and regression. The chief hypothesis here is that by developing multi-task learning models that can identify crop types and estimate yields simultaneously, we will be able to improve multi-crop yield estimations while eliminating the necessity for using CDL as model inputs.

To achieve this main objective, we defined three sub-objectives (SO) with a total of four research questions (RQ) as follows:

SO1: Assess the CDL effect on multi-crop yield estimation performance

RQ1.1: How effective in terms of accuracy is adding the CDL as a factor for estimating multi-crop yield?

SO2: Develop multi-task learning models for crop type identification and crop yield estimation.

RQ2.1: Is it feasible to use models that could be applied for segmentation such as U-net and Swin as a backbone for both segmentation and regression?

RQ2.2.: Can multi-task learning models achieve multi-crop yield accuracies comparable to CNNs with CDL as input?

SO3: Assess the performance of multi-task learning models on an unseen region and an unseen year.

RQ3.1: How accurately can multi-task learning models generalize spatially and temporally?

2. BACKBONE MODELS

In this MSc thesis, we use two models as the main backbone for developing our multi-crop yield estimations. The first is the U-net and the second is the Swin transformer. The subsequent two sections explain the architecture of both models.

2.1. U-net Architecture

U-net is a DL model designed for semantic segmentation. Its architecture, as illustrated in Figure 1, features two main paths. On the left side is the contracting path, and on the right is the expansive one. The contracting path employs the common structure of CNNs, with each step containing two 3*3 filters applied sequentially without padding. The activation function of "ReLU" follows each filter. Each step is followed by a max-pooling layer of "2*2" and a "stride=2", which is used for downsampling. Each step of downsampling doubles the channel number. In the expansive path, each step upsamples the feature maps and applies a 2*2 up-convolution that reduces the channels' number by half. Then, the resulting feature maps are concatenated with the feature maps that are cropped from the contracting path. Subsequently, two convolutions of "3*3" are applied, each followed by "ReLU". The final layer employs a 1*1 convolution to convert each 64-component feature vector to the required number of classes (Ronneberger et al., 2015).

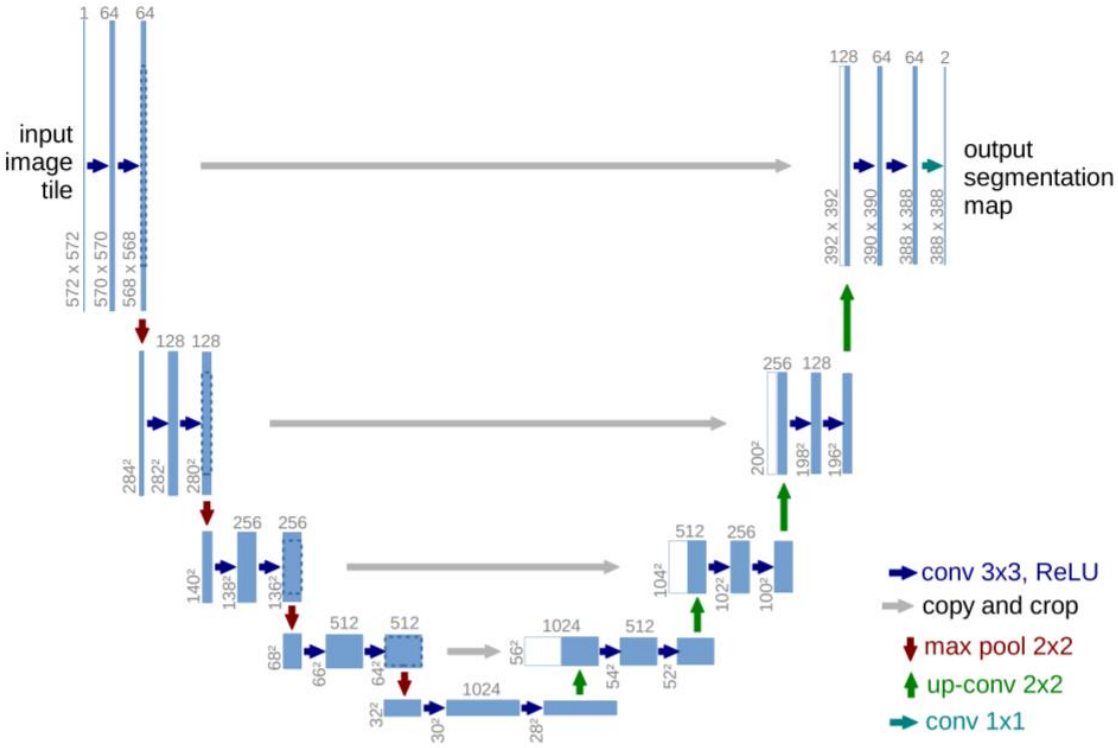


Figure 1: U-Net Architecture (Ronneberger et al., 2015))

2.2. Swin Architecture (transformer-based)

The first introduction of transformers was in 2017. It incorporated the attention mechanism in Natural Language Processing (NLP) applications (Vaswani et al., 2017). Subsequently, in 2018, The Bidirectional Encoder Representations from Transformers (BERT) model was introduced to the field of NLP (Devlin et al., 2018). It employed a pre-training mechanism on an unlabelled text within a transformer-based framework. When transformer frameworks achieved notable success in NLP, researchers began to adapt them for computer vision applications. This adaptation commenced with ViT and has since witnessed various modifications and variations.

As explained in (Dosovitskiy et al., 2020) and illustrated in Figure 2, ViT follows the architecture of the transformers applied in NLP with some modifications. The image is divided into patches in which each patch is considered a token. Those patches are then flattened from 2D matrices into 1D vectors. Those 1D vectors are reduced into lower-dimensional vectors using weight matrix multiplication and bias addition. Since all these vectors are fed into the transformer block simultaneously, they are positionally embedded so that the original image location of each patch is known to the transformer. The unique component of the transformer encoder is the multi-task learning attention, which calculates the relationships between each patch and the rest of the patches in the image. Finally, the output is used as input into a task-specific head (multi-layer perceptron), which was a classification head in the ViT paper (Dosovitskiy et al., 2020).

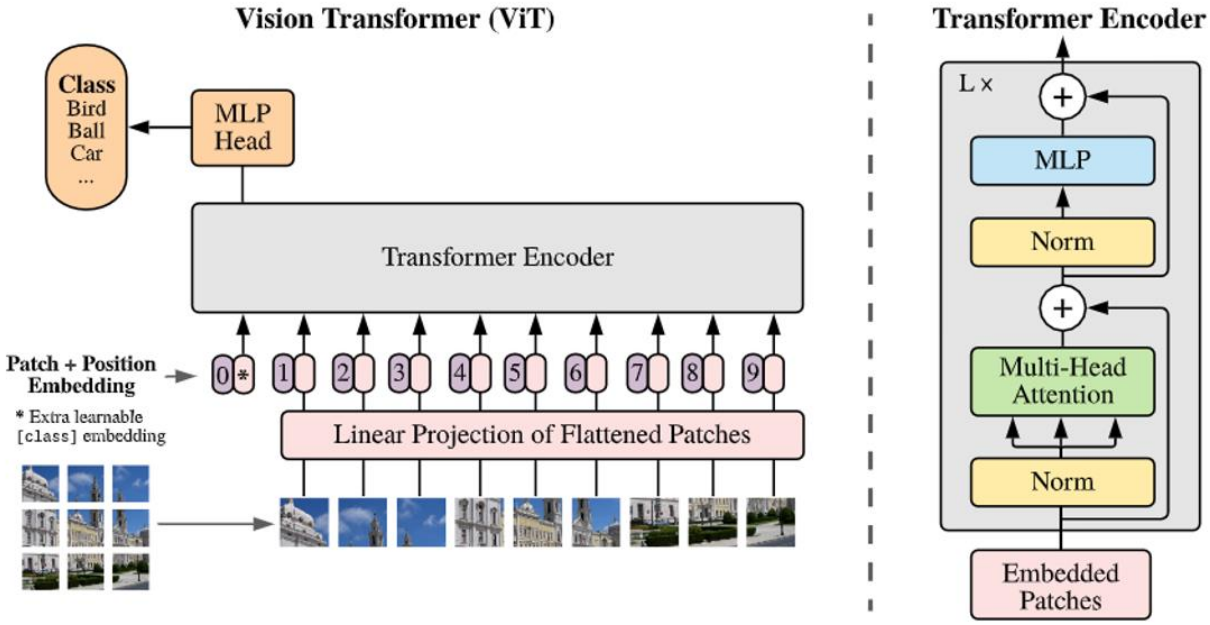


Figure 2 : The ViT architecture (Dosovitskiy et al., 2020))

On the other hand, Swin is a transformer-based model that offers hierarchical representations through a shifted window approach, and it comes with linear complexity (Liu et al., 2021). Figure 3 shows the architecture of Swin, which is acquired from the original paper. It begins with partitioning the input image into patches that do not overlap. As in ViT, every patch is considered as a token. A linear embedding is performed on each patch to transform them into 1D vectors that are understandable to the Swin transformer block to perform self-attention. However, Swin divides the image into windows and calculates the relationships between each patch and the rest only within each window. Therefore, it is called window-based "multi-head self-attention (W-MSA)" instead of only MSA in ViT. To account for the connection between the windows, it also uses "shifted-window multi-head self-attention (SW-MSA)." To gain global

information from the image, a hierarchical representation is performed using the patch merging layers that merge the adjacent patches on different stages, as depicted in Figure 3, where each block comprises patch merging and Swin transformer block. This architecture proves to be a general-purpose model in computer vision (Liu et al., 2021).

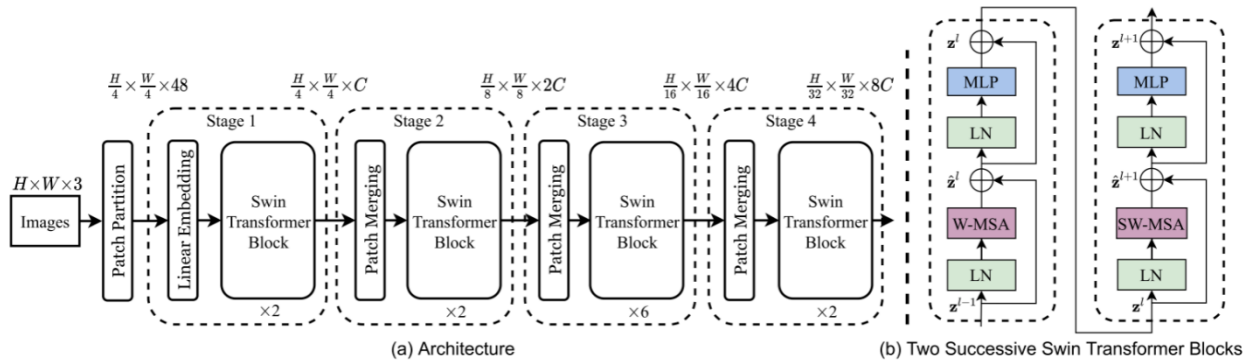


Figure 3: Swin Architecture (Liu et al., 2021)

3. DATA AND METHODS

This chapter outlines the case study and research methods used to achieve the stated research objectives. As illustrated in Figure 4, the process begins with defining the case study, followed by the data acquisition and preparation pipelines. Subsequently, we developed two categories of models: three based on CNNs and one on a transformer model (Swin). For the CNNs models, two base CNNs were developed with identical architectures but varying input channels (one incorporating CDL and one excluding them), in addition to a U-net model. To achieve sub-objective (1), the two base CNN models were compared to evaluate the impact of including CDL as an input on multi-crop yield estimation performance. Additionally, to achieve sub-objective (2), the developed multi-task learning models were compared with the CNN models to assess their performance in multi-crop yield estimation. Furthermore, to achieve sub-objective (3), we tested the multi-task learning models on an unseen region and unseen year dataset to evaluate their spatial and temporal performance.

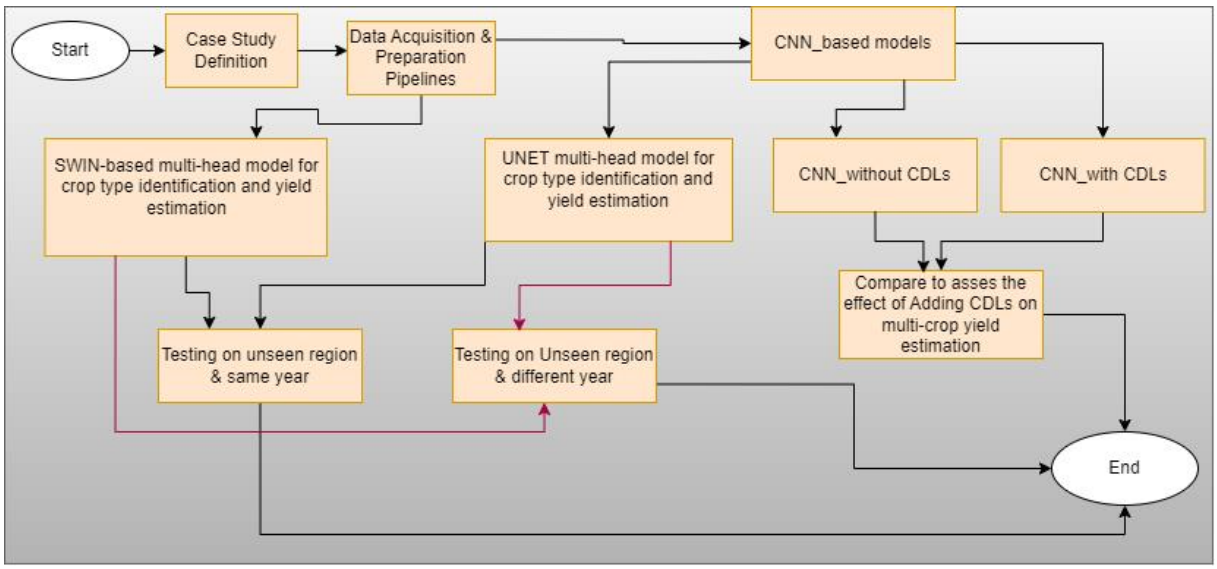


Figure 4: The research methodology

3.1. Case Study

Given the availability of crop yield data and CDL in the USA, four states are selected for our case study. The USDA provides yearly crop yield data per county. Additionally, the USDA produces annual CDL for all crop types, generated using satellite imagery and ground truth data to classify crop types (USDA. 2024). The selection criteria prioritized states with the highest corn and soybean production in the country. Consequently, the top four states in corn and soybean production are chosen, as shown in Figure 5 and Figure 6. These states are Illinois, Iowa, Minnesota, and Indiana, as depicted in Figure 7.

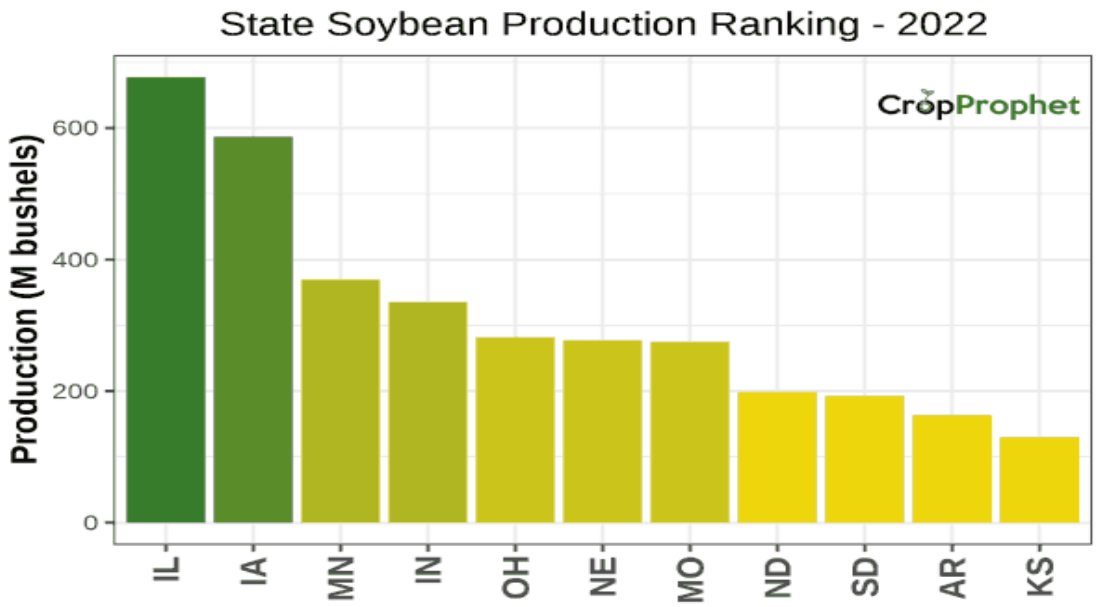


Figure 5: The state soybean production ranking in the USA in 2022 (US Soybean Production by State, 2023) Accessed on 16 June 2024

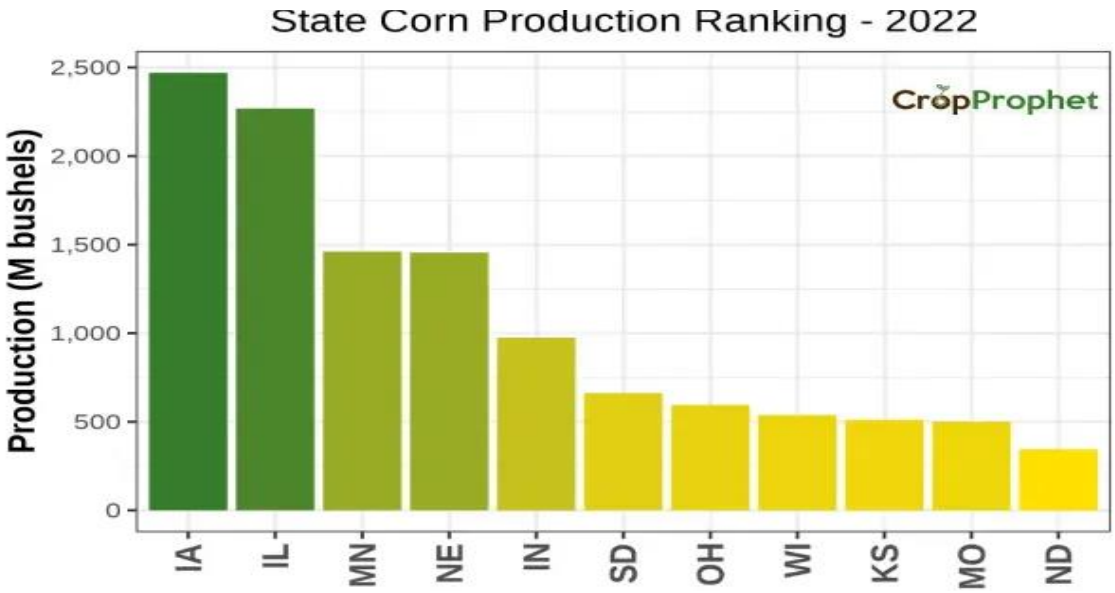


Figure 6: The state corn production ranking in the USA in 2022 (US Corn Production By State, 2023) Accessed on 16 June 2024

Figure 7: The selected four states in the USA for the research case study

3.2. Data Acquisition and Preparation Pipelines

This section explains the data acquisition and preparation pipelines. This includes the EO data (Sentinel-2), CDL, and crop yield data. USDA provides the crop yield data in BU/Acre per county. Although the works listed in Table 1, which are using USDA data utilized the crop yield data as it is and estimated the crop yield to the county level in BU/Acre, we followed a different approach. To increase the resolution at which we estimate the crop yield, we downsampled the crop yield per county into the pixel level and calculated the total yield per input image. This approach was first applied by (Venugopal, 2023). Moreover, most of the research in crop yield estimation with EO data transforms the data into histogram-like tensors of pixel intensities. This approach lacks the qualities received from the spatial dimensions of the input images (Ilyas et al., 2023). Unlike histogram tensors, we used the raw image bands as input to the developed models to make use of the spatial aspect in our estimation.

Figure 8 presents the data preparation pipeline. It was generically developed so that all the data from 2019 to 2023 could be downloaded at any time for further research, However, in this study, only data in 2022 were downloaded for training and testing, whereas data about only one state (Minnesota) in 2023 was downloaded for testing. The pipeline in Figure 8 begins with input data, including shapefiles for states and counties and a CSV file containing corn and soybean yield per county. A random points shapefile is generated within the four states of interest. These points are then intersected with the county shapefile boundaries to determine the county name for each point. Only points with available crop yield data are retained, ensuring all points have corresponding crop yield values per county, as some counties lack crop yield data). The random points shapefile is then divided by state, and each split file is replicated for different years (2019 to 2023). This approach ensures data can be prepared for future downloads for any given year from 2019 to 2023). Since the crop yield is originally measured in BU/Acre and the unit of Sentinel-2 imagery is in meters, the crop yield is converted to BU/meter using Equation 1. The crop yield CSV file is then divided by state, and each state's data is further split by year (2019 to 2023). Each of these CSV files was subsequently joined with its corresponding points shapefile. As a result, a shapefile of points is obtained for each state and year, which can be used to download both Sentinel-2 images and CDL.

Crop yield (BU/m2) = (1/4046.86) * crop yield (BU/Acre) **(1)**

Figure 9 illustrates the data downloading pipeline. It is initiated by drawing a boundary box around each random point, sized according to the intended image dimensions. These boundaries measure 224 * 224 pixels, with each pixel representing 10 meters, resulting in an image size of approximately 5 km². These boundaries are then used to download both Sentinel-2 images with eight bands (as shown in Table 2) and CDL for corn and soybean. For sentinel-2 bands, the reflectance values are averaged in the mid-season months of corn and soybean (July and August) according to the crop calendar in the USA as shown in Figure 10. In the CDL, corn is represented by a pixel value of 1, and soybean by a pixel value of 5. The CDL is further processed to reassign corn pixel values as 1, soybean pixel values as 2, and all other pixels (background) as 0. This is because the cross-entropy loss function in semantic segmentation expects pixel values to range from 0 to (number of classes - 1). Sentinel-2 images are also processed by normalizing all pixel values between 0 and 1 by dividing them by 10,000. Sentinel-2 images and CDL are saved with unique names, ensuring corresponding images share the same name. Equation 2 is utilized to calculate crop yield per image patch, and thus, each row in the final CSV files corresponds to a patch name and includes fields for the total crop yield of corn and soybean.

Y = N * CY * (r*r) (2)

Where:

- Y is Crop yield per image patch.
- N is the number of pixels of a specific crop.
- CY is crop yield (BU/m2).
- r is the image resolution (10m).

For downloading sentinel-2 images and CDL, Google Earth Engine (GEE) is utilized. Although some bands have a resolution of 20 meters and the default resolution of CDL is 30 meters, all the data downloaded from GEE is resampled to 10 meters.

Once the data are fully downloaded and processed, they are divided into training, validation, and test sets for use in the DL models. Table 2 displays the number of patches for each dataset. Three states (Iowa, Illinois, and Indiana) are used for training and validation in 2022, with 70% of the patches allocated for training and 30% for validation. The state of Minnesota in (2022 and 2023) is used for testing. Table 4 shows the average crop yield of corn and soybean per each dataset. Based on the training and validation dataset, crop yield values are normalized between 0 and 1. The minimum and maximum training yield values of corn and soybean are used to normalize the test dataset, as shown in Equations 3 and 4. This ensures that no information from the training process is leaked into the test dataset. Figure 11 presents the final output of the data acquisition and preparation pipeline, where each Sentinel-2 image patch is matched with a corresponding CDL patch and total crop yield values for corn and soybean.

Normalized Corn yield = (corn yield – minimum corn yield)/ (maximum corn yield -minimum corn yield) (3)

Normalized Soybean yield = (Soybean yield – minimum Soybean yield)/ (maximum Soybean yield -minimum Soybean yield) (4)

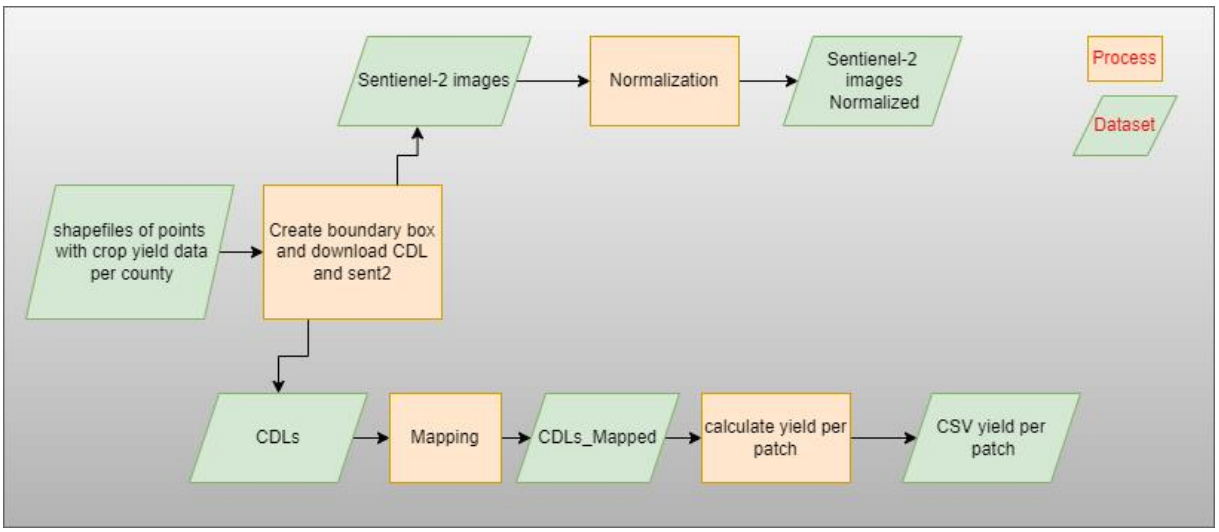


Figure 8: Data preprocessing and preparation pipeline

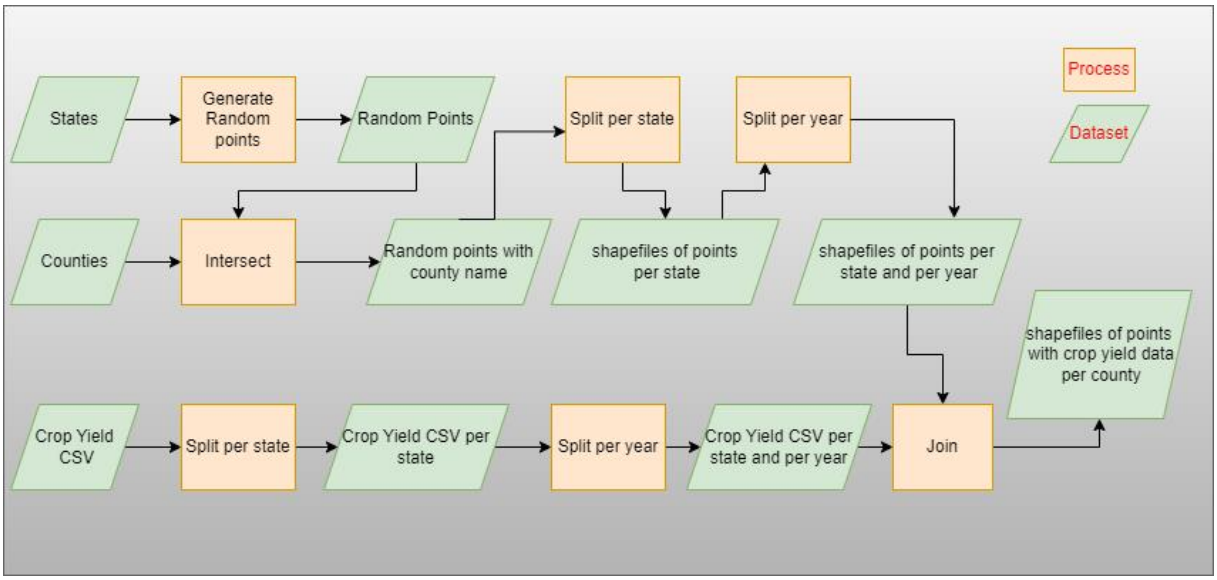


Figure 9: Data downloading and post-processing pipeline

Table 2: The selected bands of the downloaded sentinel-2 images (Acquired and modified from (Kaplan & Avdan, 2017))

Band number	Sentinel-2 Bands	Central Wavelength (Micrometre)	Resolution (m)
1	Band 2 - Blue	0.490	10
2	Band 3 - Green	0.560	10
3	Band 4 - Red	0.665	10
4	Band 5 – Vegetation red edge	0.705	20
5	Band 6 – Vegetation red edge	0.740	20
6	Band 7 – Vegetation red edge	0.783	20
7	Band 8 – NIR	0.842	10
8	Band 8A – Vegetation red edge	0.865	20

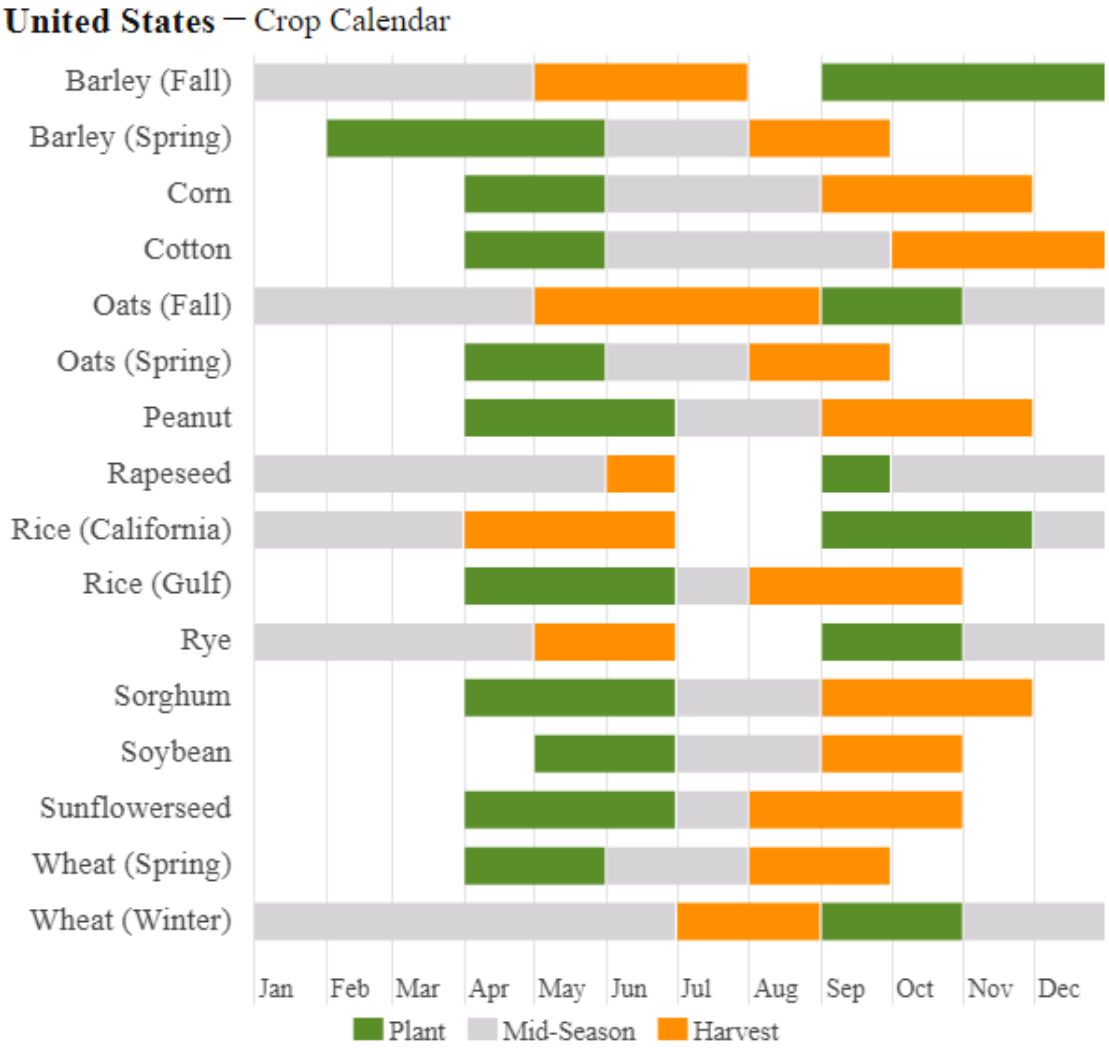
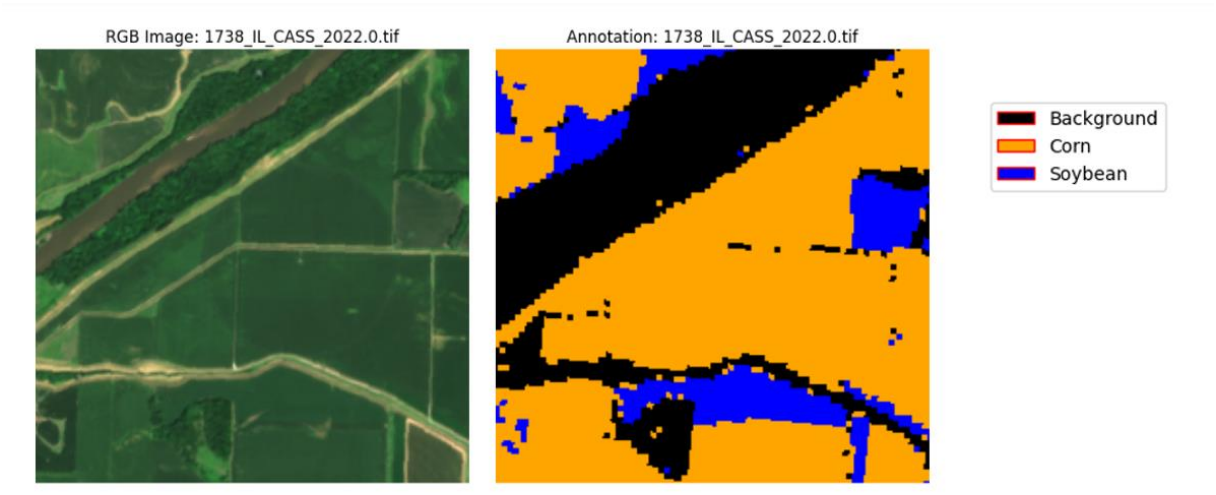


Figure 10: Crop Calendar in the USA (*United States - Crop Calendar*, 2024) Accessed on 16 June 2024



soy_y_bu_per_patch	corn_y_bu_per_patch	soy_y_normalized	corn_y_normalized
8865.846607	160784.9592	0.124577506	0.658986972

Figure 11: the final output of the data acquisition and preparation pipeline

Table 3: The number of image patches used in training, validation, and test

Training set (image patches)	Validation set (image patches)	Test set (image patches)	
5446	2338	2872	2258
Indiana, Iowa, Illinois in 2022		Minnesota in 2022	Minnesota in 2023

Table 4: The average crop yield of training, validation, and test datasets in BU

	Average corn yield per image patch (BU)	Average soybean yield per image patch (BU)
Training dataset	78730	21458
Validation dataset	79027	21788
Minnesota 2022 test dataset	54064	12249
Minnesota 2023 test dataset	59666	12760

3.3. Models' Training and Evaluation Configurations

Common loss functions and evaluation metrics were employed in all the developed models, which are explained in the following sections. As depicted in Table 5 below, Cross-entropy loss was used for segmentation tasks, whereas Mean Squared Error (MSE) loss was utilized for regression tasks. Furthermore, Intersection Over Union (IOU) was used to evaluate segmentation tasks, while RMSE and the coefficient of determination (R^2) metrics were used to evaluate the regression tasks. Due to limited time, all models were only run for 50 epochs with Adam optimizer, and the learning rate scheduler was chosen as "ReduceLROnPlateau.". The Early Stopping was configured with (Patience = 10). All DL models and the loss functions were implemented using PyTorch. Table 6 explains the concepts of all the models' configurations mentioned in Table 5.

The developed models are multi-headed; thus, each head has a loss function. However, a single loss function is required for backpropagation. A loss combination method was used to achieve this. In this research, we used the total sum of losses for backpropagation. Equation 5 illustrates the combined loss function of the multi-head regression CNN models for estimating corn and soybean yields, while Equation 6 was used to calculate the combined losses with the multi-head regression and segmentation models (U-net and Swin).

Combined loss (two-head regression model) = loss1 + loss2 (5)

Where:

- loss1: the loss of the first head calculated using MSE.
- loss2: the loss of the second head calculated using MSE.

Combined loss (multi-head segmentation and regression model) = loss1 + loss2 + loss3 (6)

Where:

- loss1: the loss of the first head calculated using MSE.
- loss2: the loss of the second head calculated using MSE.
- Loss3: the loss of the segmentation head calculated using Cross Entropy

Table 5: DL models' common configurations used across the research

	Regression Tasks	**Segmentation Tasks**
Loss Function	MSE	Cross-Entropy
Evaluation Metrics	MSE, R^2	IOU
Optimizer	Adam	
Epochs	50	
Learning Rate (LR)	0.0001	
LR Scheduler	PeduceLROnPlateau	
Early Stopping	(Patience = 10)	

Table 6: Explanation of all the DL models' parameters used in the research

MSE (Mean Squared Error)	Is utilized as a loss function for regression problems. It is calculated as the average of the squared residuals, where residuals are the difference between true and predicted values (Jadon et al., 2022).
R^2	Is a metric used to assess regression models, indicating how well the predicted values match the target values. It quantifies the extent to which the independent variables account for the variation in the dependent variable (Tatachar, 2021).
RMSE	Is an evaluation metric of regression problems that defines how close the outcomes from the model are to the label values (Tatachar, 2021).
Intersection over Union (IOU), or Jaccard Similarity Index (JSI)	Is a metric used for evaluating segmentation problems, calculated by determining the ratio of the overlapping area between the predicted segmented map and the ground truth label map. (Rizwan I Haque & Neubert, 2020).
Cross-entropy (CE)	is a statistical measure utilized to assess the disparity between two probability distributions associated with a specific random variable. This metric is particularly advantageous in numerous machine learning applications, such as semantic segmentation and classification tasks. Within the field of semantic segmentation, cross-entropy loss quantifies the extent to which a model's predictions correspond with the actual target labels (Azad et al., 2023).
Adam	is an algorithm based on gradients to optimize stochastic objective functions. It is robust for large

	data problems and memory-efficient (Kingma & Ba, 2014).
ReduceLROnPlateau	is a dynamic learning rate technique that reduces the learning rate when the validation loss stops improving for a specific number of epochs (*ReduceLROnPlateau — PyTorch 2.3 Documentation*, 2024).

3.4. Models' Architectures

To address the identified research gaps and to achieve the stated research objectives, we developed three models. The first model was based on CNN. It was developed to assess the impact of adding CDL as extra input with sentinel-2 images on multi-crop yield estimation. Therefore, the CNN model is used once with only sentinel-2 images and another time with including CDL. The other two models are developed to evaluate the applicability and performance of multi-task learning models in accurately estimating multi-crop yields. These two models identify the crop type and estimate its yield concurrently. Therefore, two tasks are implemented simultaneously (segmentation and regression). The two multi-task learning models are based on backbone architectures that work for segmentation. The first utilized the U-net architecture and the second employed the Swin architecture.

3.4.1. CNN architecture

This model is based on CNNs. It comprises several layers, including convolutional layers, max-pooling layers, fully connected layers, batch normalization, dropout, and two output heads. The main objective of this model is to concurrently estimate the crop yield of both corn and soybean. The architecture, as depicted in Figure 12, is applied in two scenarios. The first scenario uses only the eight bands of Sentinel-2 as input. The second scenario incorporates the eight bands of Sentinel-2 along with CDL as the ninth band. Since all Sentinel-2 bands are normalized from 0 to 1 in the second scenario, the CDL pixel values are also remapped so that 0 represents the background, 0.5 represents corn, and 1 represents soybean. Figure 12 illustrates the model's architecture. It includes 4 2D convolutional layers with (kernel size = 3), each followed by a max-pooling layer with (kernel size = 2). After flattening the output from the final max-pooling layer, four fully

connected layers are applied. Subsequently, two separate heads are created: one for corn yield estimation and the other for soybean yield estimation.

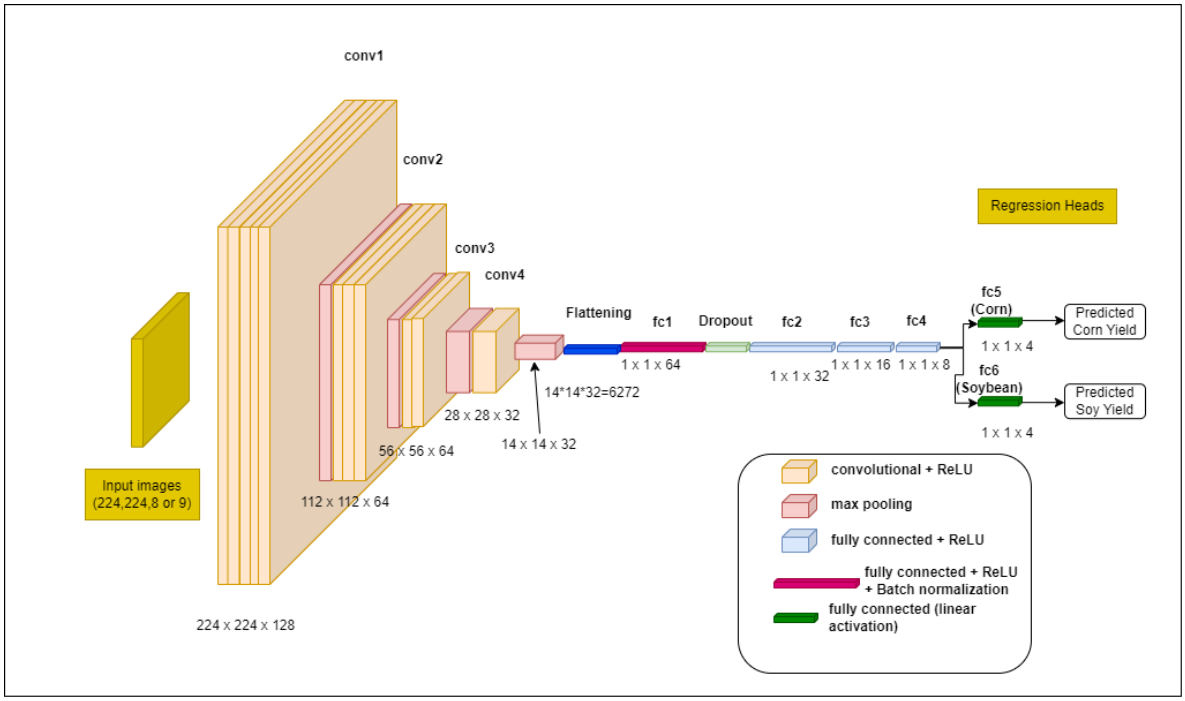


Figure 12: The architecture of the two CNNs models (with CDL and without CDL)

3.4.2. UNET architecture

This model, as detailed in Figure 13, is a modified version of the U-Net architecture. It follows the Shared Trunk approach of multi-task learning models where the U-net is the employed backbone for feature extraction. The model is developed for segmentation and yield estimation, featuring an encoder-decoder structure. The encoder path comprises three convolutional blocks with ReLU activations, followed by max-pooling layers that reduce the spatial dimensions and increase the feature channels. The bottleneck layer further processes the features with two convolutional layers. The decoder path mirrors the encoder. It implements upsampling through transposed convolutions and concatenation with corresponding encoder features. The third step in the decoder path is branched into two heads, the segmentation head to restore the original spatial dimensions (segmentation maps) while the second head is for regression. The first part of the regression head includes the Adaptive Average Pooling (AAP) process. The pooled output is then flattened. Subsequently, two separate heads are created: one for corn yield estimation and the other for soybean yield estimation.

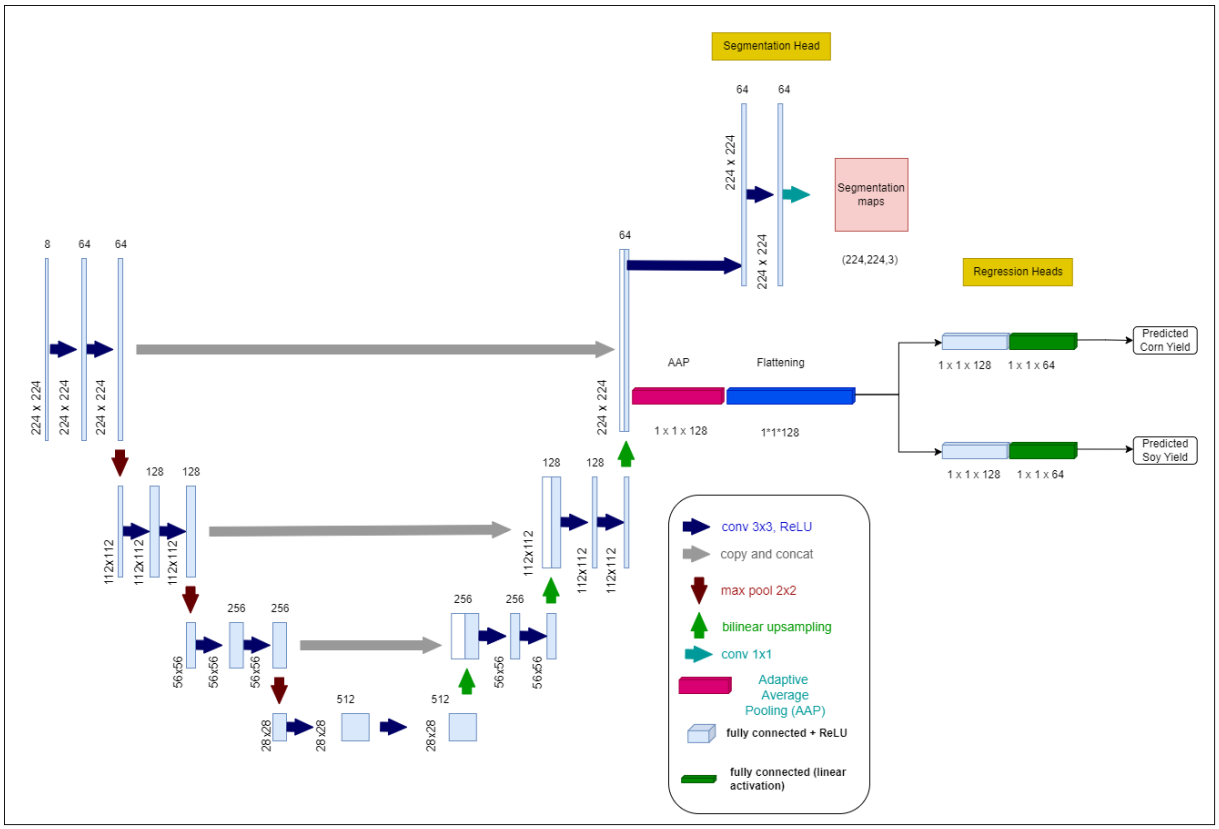


Figure 13: The architecture of the multi-task learning U-Net model

3.4.3. SWIN-based model

This model explained in Figure 14, is utilizing the Swin model architecture as a feature extraction backbone following the Shared Trunk approach of multi-task learning models. It is developed for crop type identification and yield estimation, featuring an encoder-decoder structure. As explained before, in Figure 3, each of the four stages in the encoder (Swin) increases the channels' number while reducing the spatial dimensions. Two separate heads are designed. The segmentation head consists of four transposed convolutional layers that upsample the feature maps that are then concatenated with lower-stage feature maps to refine the segmentation output (original image size). The regression head includes adaptive average pooling on the output feature maps from Swin, which are then concatenated. A series of fully connected layers process the concatenated features and separately predict the outputs for corn and soybean using distinct regression heads for each crop type.

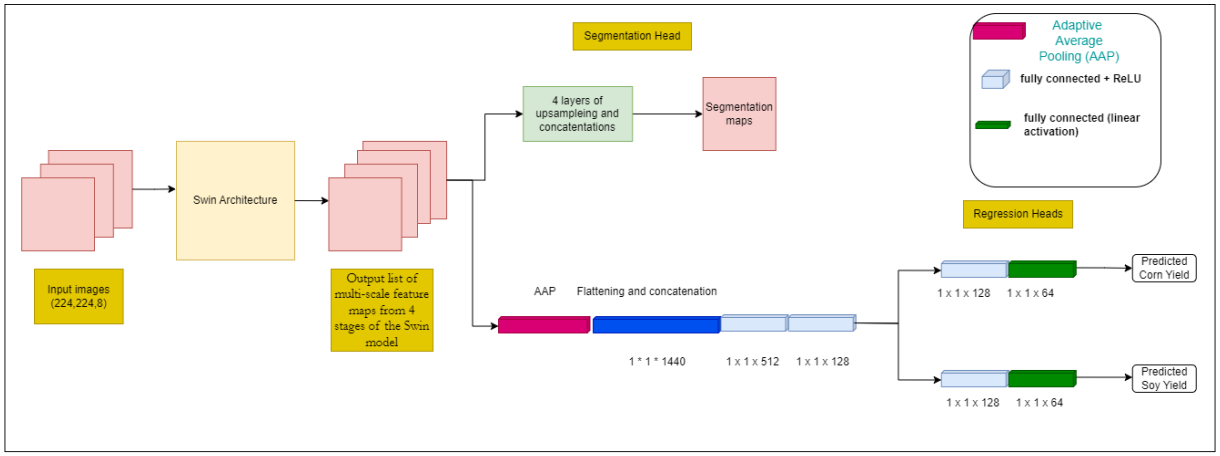


Figure 14: The architecture of the multi-task learning Swin model

3.5. Models' Testing and Comparisons

The testing phase is used to assess the developed models on unseen data. As shown above in Table 3, section 3.2, two datasets were prepared for testing purposes Minnesota 2022 and Minnesota 2023. Both Minnesota 2022 and Minnesota 2023 are in different regions from the regions the models were trained on. Moreover, Minnesota 2023 is in a different region and different year knowing that all the models were trained on data in 2022 in three states (Iowa, Indiana, and Illinois). For testing, the two CNN models were tested on the Minnesota 2022 test dataset whereas the U-net and Swin models were tested on both The Minnesota 2022 and Minnesota 2023 test datasets.

When testing, the predicted crop yield values were first denormalized to the BU unit for all the models. Then, the evaluation metrics for regression (RMSE and R^2) were calculated. However, the predicted values were negative for some patches that contained almost zero yield. Therefore, they were transformed to 0 before the denormalization process.

To answer the research questions stated in section 1.4., we needed to do comparisons on the output from the developed models.

Three comparisons are implemented:

1- For question 1, CNN model with CDL and CNN model without CDL to assess the impact of adding CDL as input on multi-crop yield estimation.
2- For questions 2 and 3, the CNN model with CDL and the two multi-task learning models (U-Net and Swin) to assess the performance of multi-task learning models in estimating multi-crop yield.
3- For question 4, Testing (U-Net and Swin) multi-task learning models on an unseen region and unseen year to assess their performance spatially and temporarily on estimating the yield of multi-crops.

3.6. Code and Reproducibility

This section provides an overview of the implementation part of all models. It highlights the libraries and computational resources utilized. Moreover, it explains the code structure on the thesis GitHub repository.

PyTorch, which is a library for developing and training ML and DL models, was used for all the developed models. All the CNN-based models were straightforward to implement using PyTorch built-in functions.

However, for the Swin model, the official implementation was based on the MMsegmentation library. MMsegmentation is a framework for unified implementations of semantic segmentation algorithms. It is a part of the OpenMMLab project for computer vision algorithms (MMSegmentation 1.2.2 Documentation, 2024). However, MMsegmentation is only designed for semantic segmentation and does not include any regression implementations. Moreover, it was by default designed to use the PIL library that does not read ".tif" images. However, in our work, we developed the images' reading functions based on the "Rasterio" library to make use of all the bands in the EO data.
A substantial time was spent on modifying the main framework of MMsegmentation to accommodate multi-task learning models for segmentation and regression. We modified the library to read satellite images, the corresponding CDL and the crop yield data. However, the modification of the training, validation and test built-in functions for regression and segmentation required significant time. Therefore, we developed separate codes for the remaining DL processes of training, validation, and testing. Additionally, specialized data loading functions were developed to read the sentinel-2 images and the corresponding CDL and corn and soybean yield values.

The main advantage of using MMsegmentation is its availability of many segmentation models. Thus, with the ready codes of various segmentation models, future work could easily experiment with different backbones for multi-task learning models using our developed framework for training, validation, and testing. However, in this research, we utilized only the Swin architecture from MMsegmentation.

For the computational resources, I first used my machine with an NVIDIA T600 Laptop GPU and 32 GB of RAM on a small sample of the data to develop the models. Running one model took from five to six hours. This slowed progress due to the development of multiple models and different comparisons. But later I had access to the Shaken server which features a 24 GB NVIDIA RTX A5000 GPU and 1.5 TB of RAM. Utilizing up to 11 GB of GPU reduced the training time of each model to around two to three hours. This helped me to speed up the process of developing the models' architectures on the full data. Although I had also access to a 64 GB GPU server, it was not always free to use as it had multiple users at the same time. However, I used it for many experiments. In the end, all the final models were run on the Shaken server.
All codes are available on the GitHub Repository. The files are organized in Jupyter Notebooks and ".py" files. They are categorized into folders such as ("DataPreparationModelBuilder," "DataDownloadingAndProcessing," "PrepareForDLModels," and "DL_Models"). For all the models, the training and validation log files and the output checkpoints at every epoch were saved.

4. RESULTS

This chapter outlines the results of all the developed models (CNN without CDL, CNN with CDL, U-net, and Swin). For each model, the following graphs are included:

1- Training and validation losses (for both regression and segmentation tasks)
2- Evaluation metrics on validation data (for both regression and segmentation tasks)
3- Evaluation metrics on test data

4.1. CNN-based models results

4.1.1. Results of the CNN model without CDL

This section explains the results of the CNN model without including CDL as input. Figure 15 illustrates the MSE losses for training and validation, along with the R^2 values, for corn and soybean, with corn depicted on the left side and soybean on the right. The training learning curves consistently decreased as the number of epochs increased, whereas the validation learning curves exhibited fluctuations up to epoch 20 and then flattened. The early stopping was triggered at epoch 37. However, the optimal model for the test dataset was identified at epoch 20 where the model flattened.
When evaluated on the Minnesota 2022 dataset, as shown in Figure 16 and Figure 17, the predicted values' distribution diverged from that of the target values. Specifically, the predicted crop yield values for corn and soybeans showed significant deviation from the 1:1 line, with soybean results being less accurate than those for corn. The data range of predicted and target yield values slightly differs for both corn and soybean. Detailed results on the Minnesota 2022 test dataset, including RMSE and R^2 values, are provided in Table 7. Additionally, Table 7 includes the average crop yield per patch for corn and soybean as 54064 BU and 12249 BU, respectively. This is used to indicate how much different the RMSE is from the average yield value per crop type.

Table 7: Evaluation metrics of the CNN model without CDL on the Minnesota 2022 test dataset

	Minnesota 2022	
	Corn	Soybean
RMSE	24456.7 BU	7724 BU
R^2	0.778	0.550
Average crop yield/patch	54064 BU	12249 BU

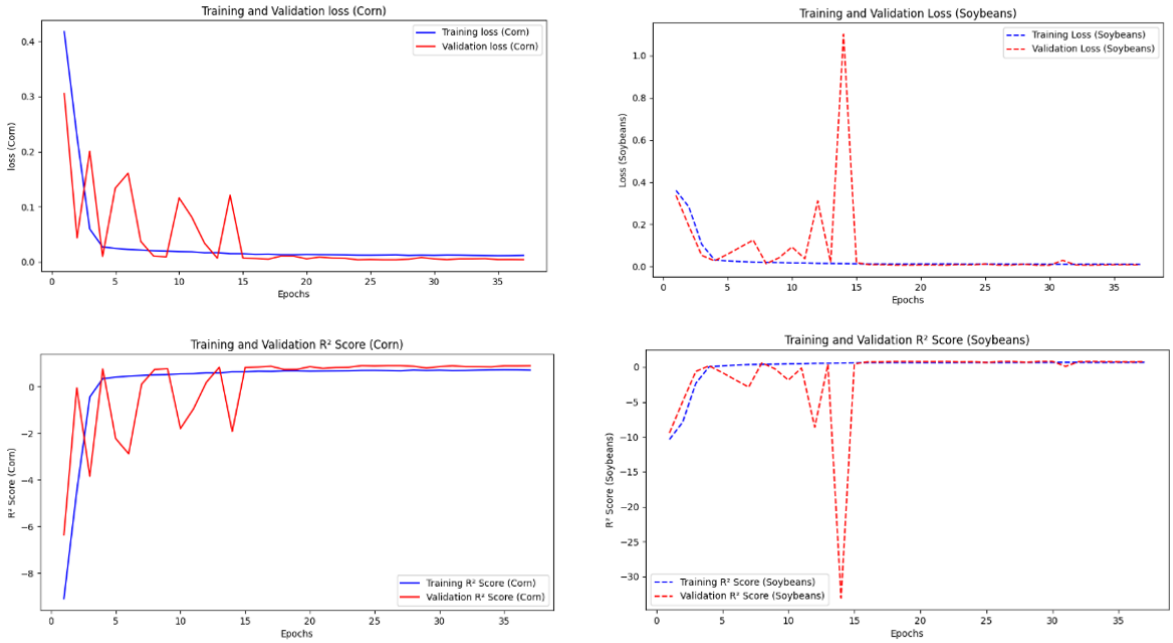


Figure 15: Training and validation learning curves of CNN model without CDL for corn (left) and soybean (right)

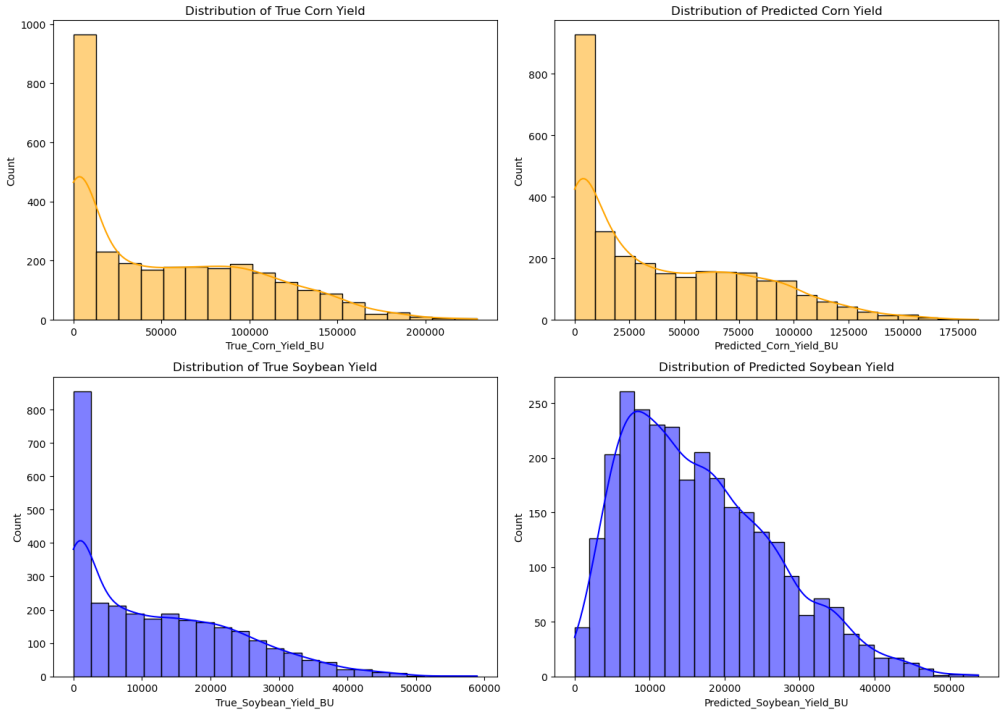


Figure 16: Distribution of corn and soybean yields of True values (left) and Predicted values (right) on Minnesota 2022 test dataset using the CNN model without CDL

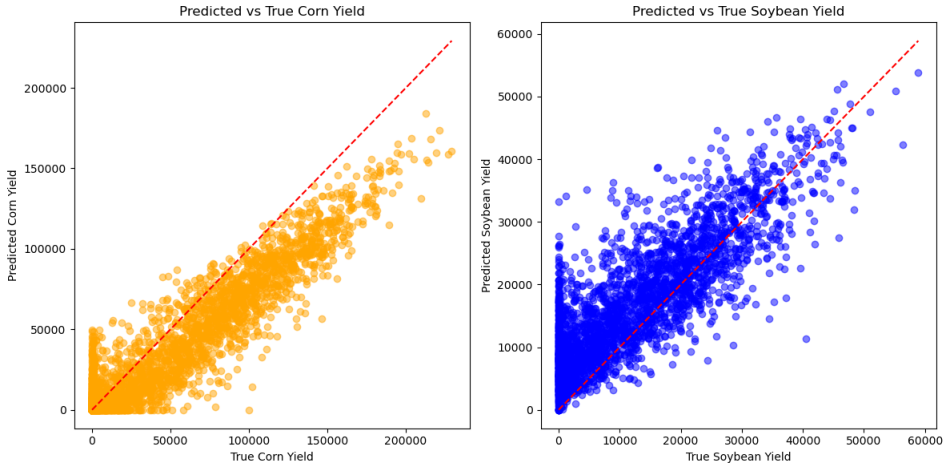


Figure 17: The 1:1 line of true and predicted yields of corn (left) and soybeans (right) on Minnesota 2022 test dataset using the CNN model without CDL

4.1.2. Results of the CNN model with CDL

This section explains the results of the CNN model with CDL fed to the model as an input. Figure 18 illustrates the MSE losses for training and validation, along with the R^2 values, for corn and soybean, with corn depicted on the left side and soybean on the right. The training and validation learning curves consistently decreased as the number of epochs increased. Although the model converged at epoch 32 for corn training and validation curves, there were still some fluctuations for soybean curves. Consequently, the optimal model for the test dataset was identified at the last epoch of 50 since the model was still learning and needed more epochs. When evaluated on the Minnesota 2022 dataset, as shown in Figure 19 and Figure 20, the predicted values' distribution was close to that of the target values. Specifically, the predicted crop yield values for both corn and soybean showed a good alignment with the 1:1 line, with corn results being slightly more accurate than those for soybean. The data range of predicted and target yield values is very close. Detailed results, including RMSE and R^2 values, are provided in Table 8. Table 8 also included the average crop yield per patch for corn and soybean as 54064 BU and 12249 BU, respectively. This is used to indicate how much different the RMSE is from the average yield value per crop type.

Table 8: Evaluation metrics of the CNN model with CDL on the Minnesota 2022 test dataset

	Minnesota 2022	
	Corn	Soybean
RMSE	10468 BU	3256 BU
R^2	0.959	0.923
Average crop yield/patch	54064 BUy	12249 BU

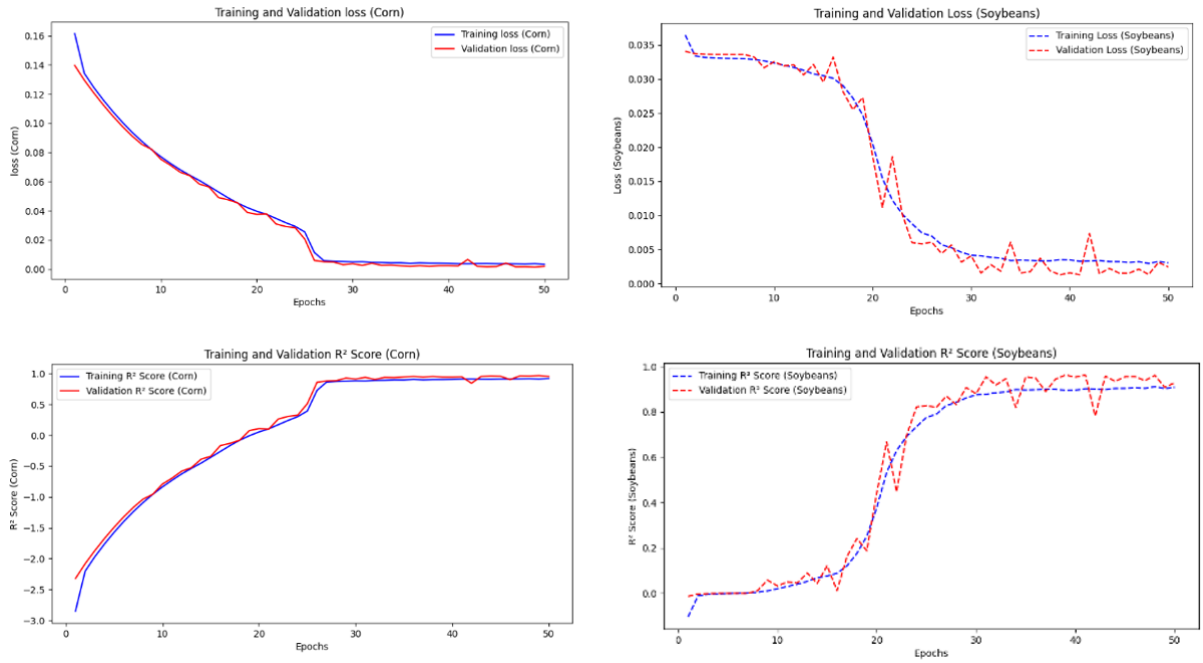


Figure 18: Training and validation learning curves of CNN model with CDLs for corn (left) and soybean (right)

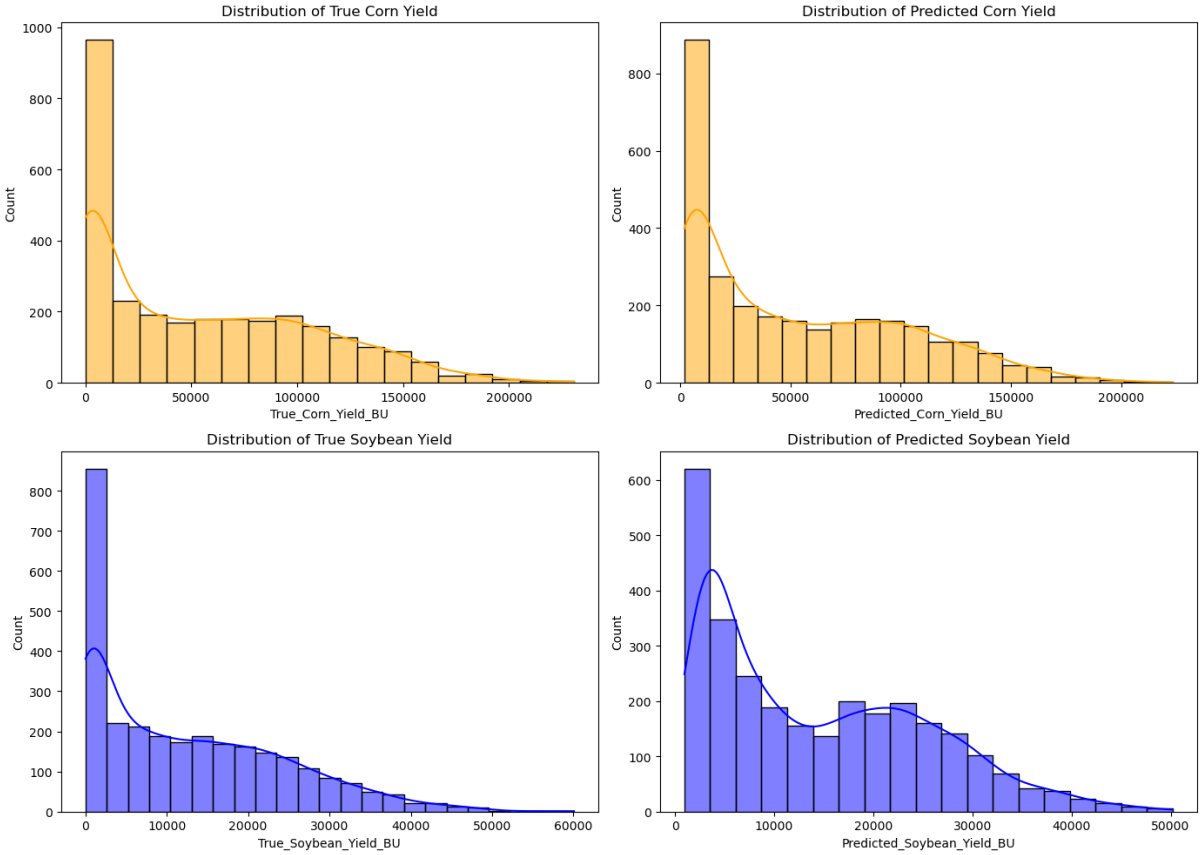


Figure 19: Distribution of corn and soybean yields of True values (left) and Predicted values (right) on Minnesota 2022 test dataset using the CNN model with CDL

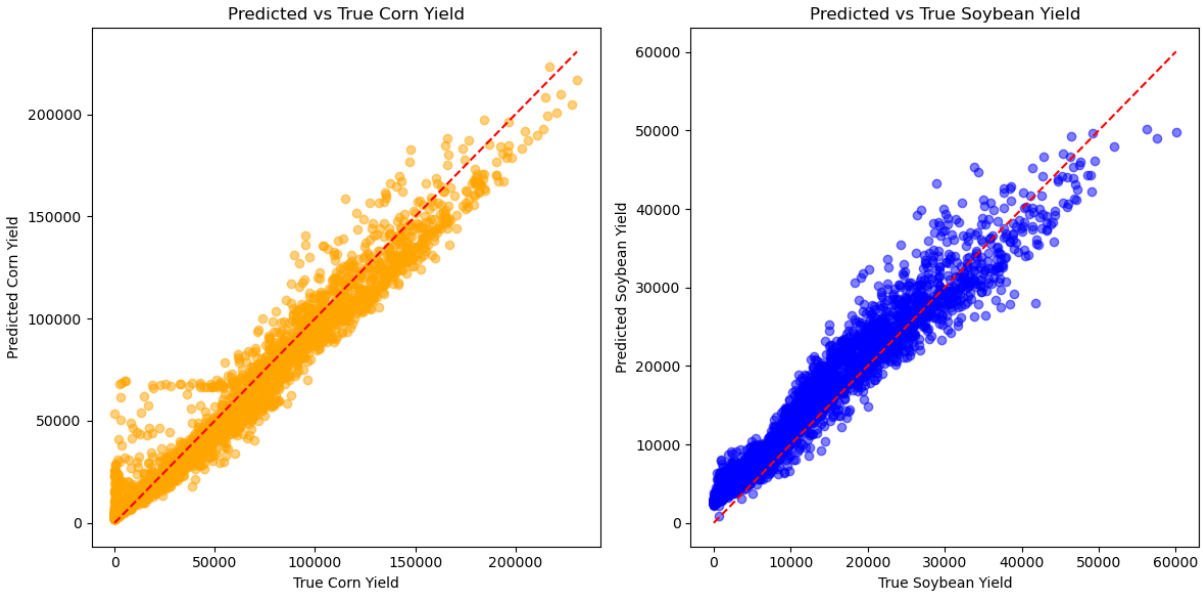


Figure 20: The 1:1 line of true and predicted yields of corn (left) and soybeans (right) on Minnesota 2022 test dataset using the CNN model with CDL

4.1.3. U-net

This section explains the results of the U-net multi-task learning model on both the Minnesota 2022 and Minnesota 2023 test datasets to assess the spatial and temporal generalizability of the model. Figure 21 presents the learning curves of the multi-task learning U-net model, emphasizing the regression loss measured by MSE and the R^2 regression metric and the segmentation loss measured by cross-entropy and the IOU segmentation metric. As the epochs progressed, the regression and segmentation losses exhibited a steady decline, while the R^2 and IOU metrics consistently increased. The model's performance almost flattened around epoch 50. Consequently, the model from epoch 50 was tested using data from Minnesota 2022 and Minnesota 2023.

Figure 22 and Figure 23 display the distribution and the 1:1 line of the target and the predicted data respectively based on the Minnesota 2022 test dataset. The figures show that the distribution of the predicted corn yield is more identical to the target values than the soybean. Furthermore, both are close to the 1:1 line with corn aligned more.

Figure 24 and Figure 25 present the distribution and the 1:1 line of the target and the predicted data respectively based on the Minnesota 2022 test dataset. The figures show that the distribution of the predicted corn yield and soybean yield differed, meaning the accuracy in the future data (Minnesota 2023) dropped compared to the current year data (Minnesota 2022). However, both are still close to the 1:1 line with corn aligned more.

From Figures 22, 23, 24, 25 and Table 9, the model's accuracies for segmentation (IOU) and regression on the Minnesota 2022 test dataset were better than those of Minnesota 2023. Furthermore, both cases had similar data ranges between predicted and target crop yield values of corn and soybean. Table 9 lists the exact values of (RMSE, R^2, IOU), of the testing process.

Table 9: Evaluation metrics of the multi-task learning U-net model in Minnesota 2022 and Minnesota 2023

	Minnesota 2022		Minnesota 2023	
	Corn	Soybean	Corn	Soybean
RMSE	14595 BU	5889 BU	19304 BU	5924 BU
R^2	0.921	0.738	0.836	0.652
Average crop yield/patch	54064 BU	12249 BU	59666 BU	12760 BU
IOU	0.7734		0.7194	

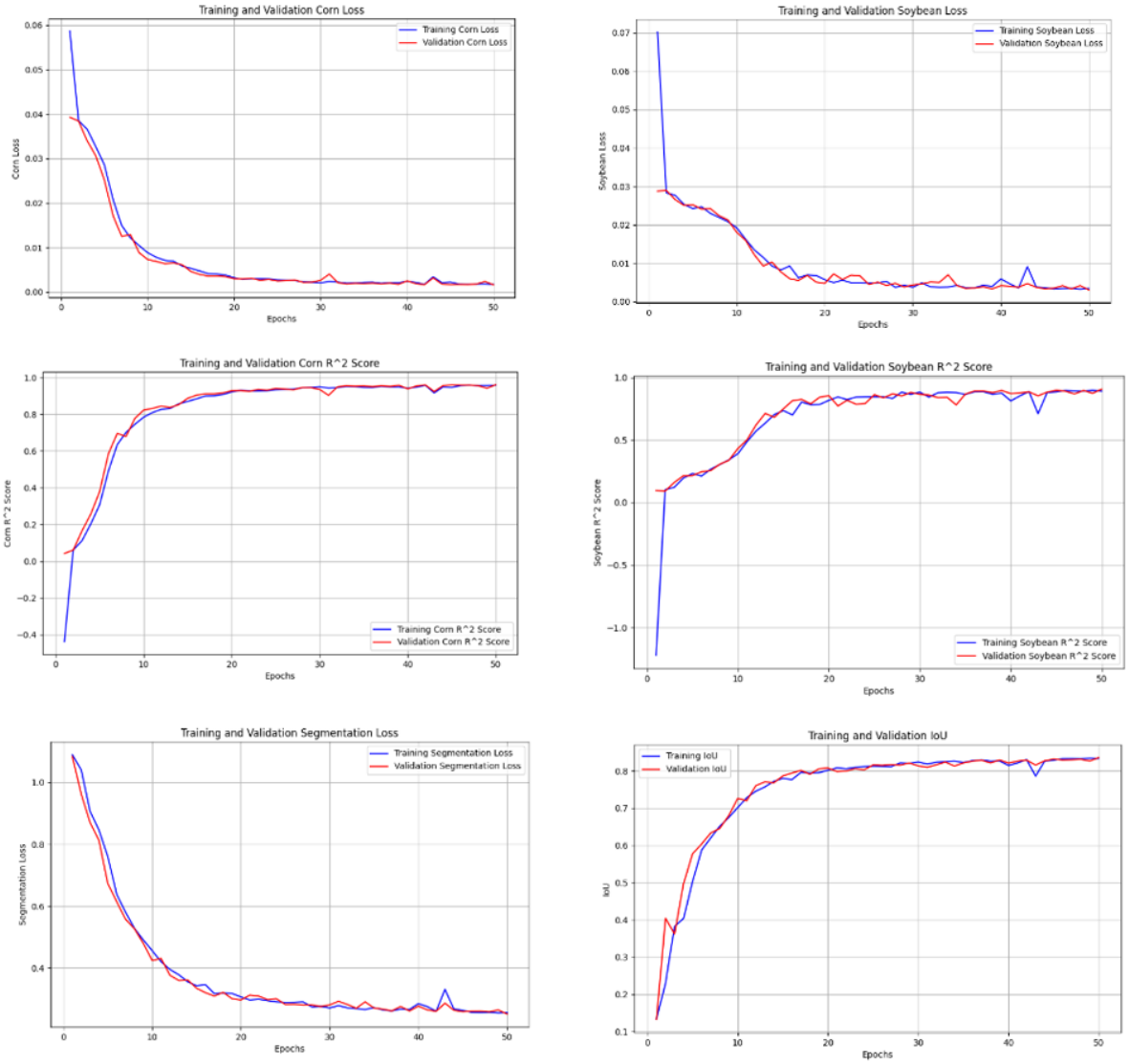


Figure 21: Training and validation learning curves of the multi-task learning U-net model (regression and segmentation)

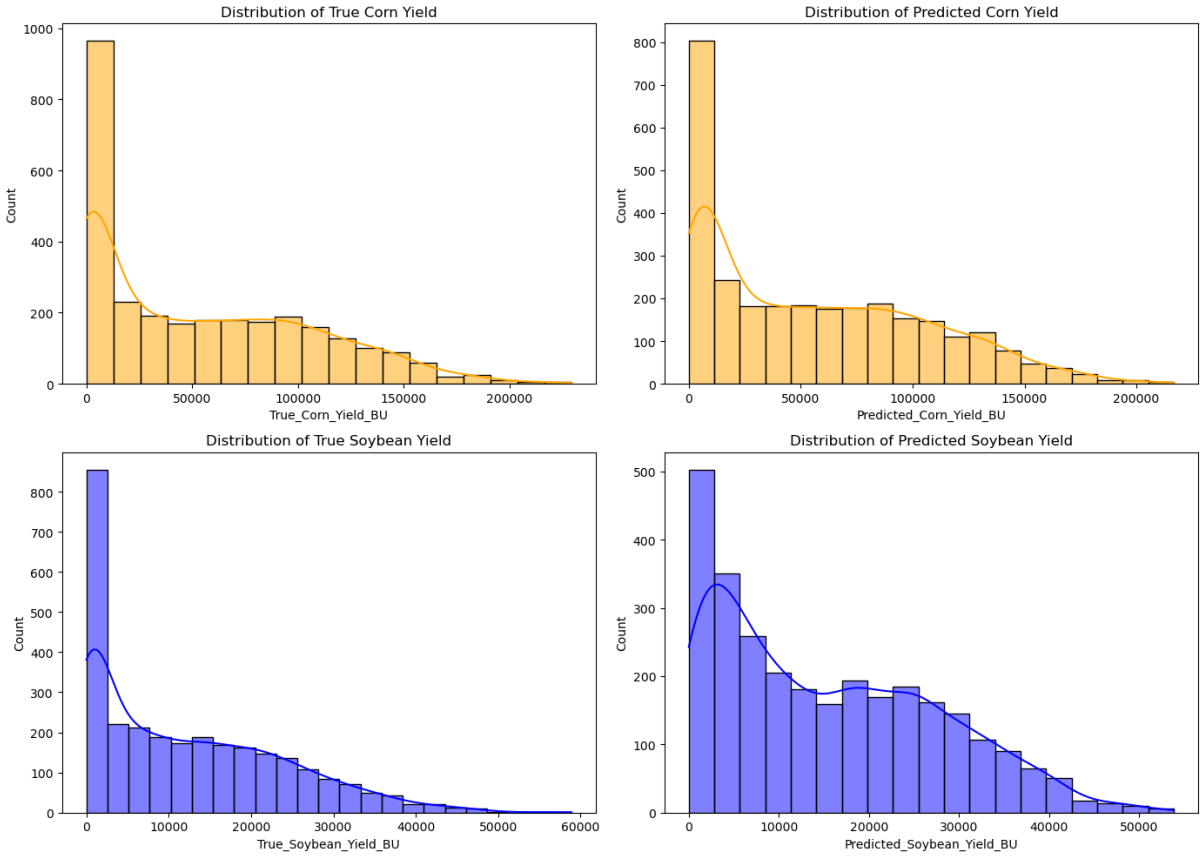


Figure 22: Distribution of corn and soybean yields of True values (left) and Predicted values (right), Minnesota-22 using the U-net multi-task learning model

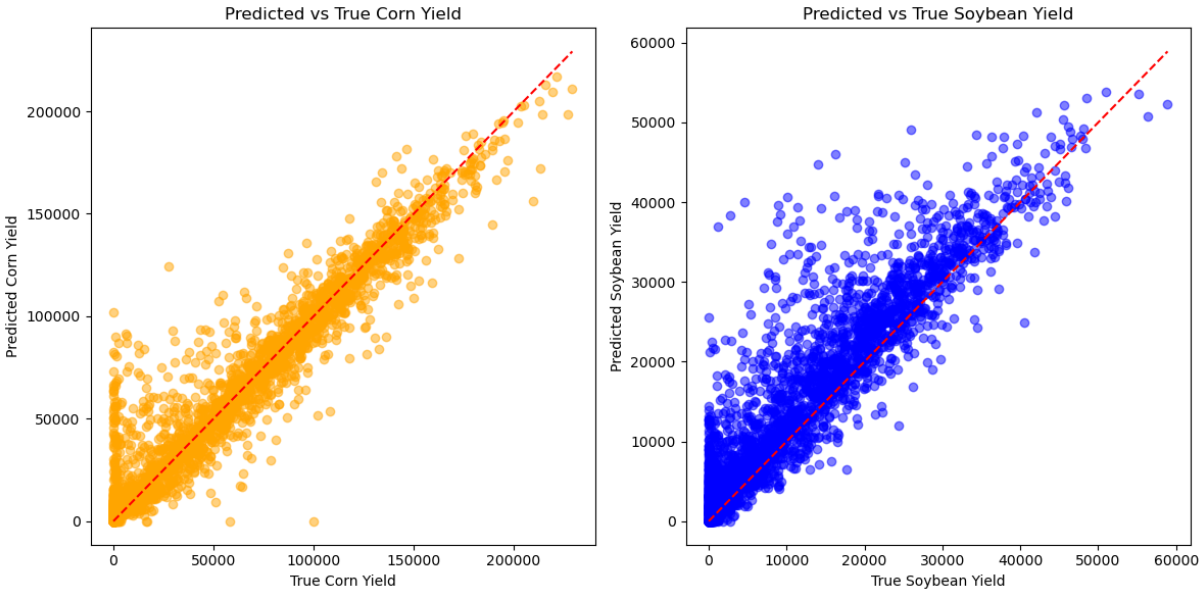


Figure 23: The 1:1 line of true and predicted yields of corn (left) and soybeans (right), Minnesota-22 using the U-net multi-task learning model

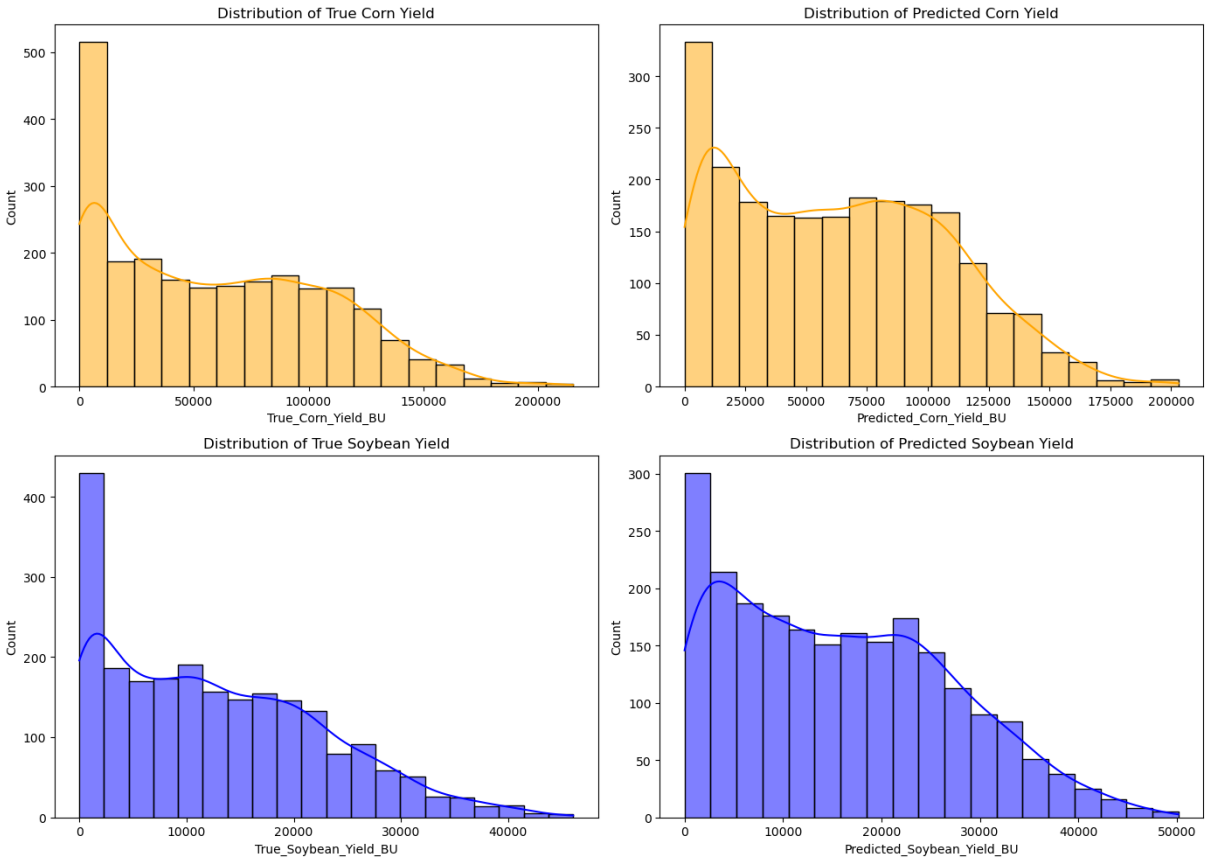


Figure 24: Distribution of corn and soybean yields of True values (left) and Predicted values (right), Minnesota-23 using the U-net multi-task learning model

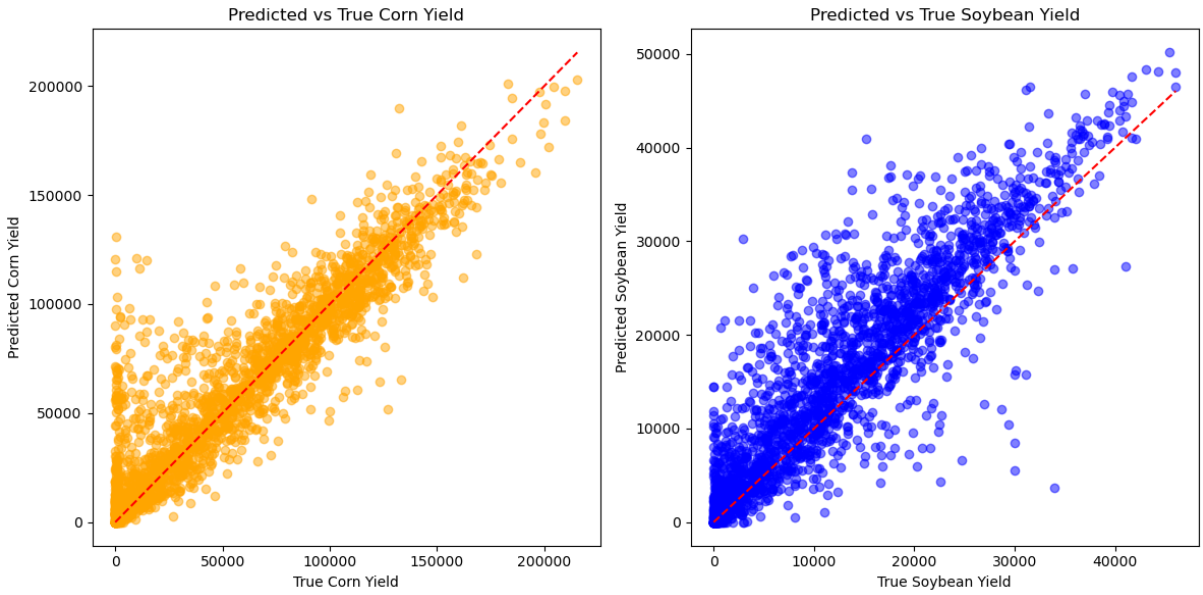


Figure 25: The 1:1 line of true and predicted yields of corn (left) and soybeans (right), Minnesota-23 using the U-net multi-task learning model

4.2. Swin model's results

This section explains the results of the Swin multi-task learning model on both the Minnesota 2022 and Minnesota 2023 test datasets to assess the spatial and temporal generalizability of the model. Figure 26 presents the learning curves of the multi-task learning Swin model, emphasizing the regression loss measured by MSE and the R^2 regression metric, and the segmentation loss measured by cross-entropy and the IOU segmentation metric. As the epochs progressed, the regression and segmentation losses exhibited a steady decline, while the R^2 and IOU metrics showed a consistent increase. The model's performance plateaued around epoch 40. Consequently, the model from epoch 40 was tested using data from Minnesota 2022 and Minnesota 2023.

Figure 27 and Figure 28 display the distribution and the 1:1 line of the target and the predicted data respectively based on the Minnesota 2022 test dataset. The figures show that the distribution of the predicted corn yield is more like the target values than the soybean. Moreover, both are close to the 1:1 line with corn aligned more.

Figure 29 and Figure 30 present the distribution and the 1:1 line of the target and the predicted data respectively based on the Minnesota 2022 test dataset. The figures show that the distribution of the predicted corn yield and soybean yield differed, meaning the accuracy in the future data (Minnesota 2023) dropped compared to the current year data (Minnesota 2022). However, both are still close to the 1:1 line with corn aligned more.

From Figures 27, 28, 29, and 30, and Table 10, the model's accuracies for segmentation (IOU) and regression on the Minnesota 2022 test dataset were better than those of Minnesota 2023. Furthermore, both cases had similar data ranges between predicted and target crop yield values of corn and soybean. Table 10 lists the exact values of (RMSE, R^2, IOU), of the testing process.

Table 10: Evaluation metrics of the multi-task learning Swin model in Minnesota 2022 and Minnesota 2023

	Minnesota 2022		Minnesota 2023	
	Corn	Soybean	Corn	Soybean
RMSE	15493 BU	5047 BU	21842 BU	5375 BU
R^2	0.911	0.808	0.79	0.714
Average crop yield/patch	54064 BU	12249 BU	59666 BU	12760 BU
IOU	0.8001		0.7611	

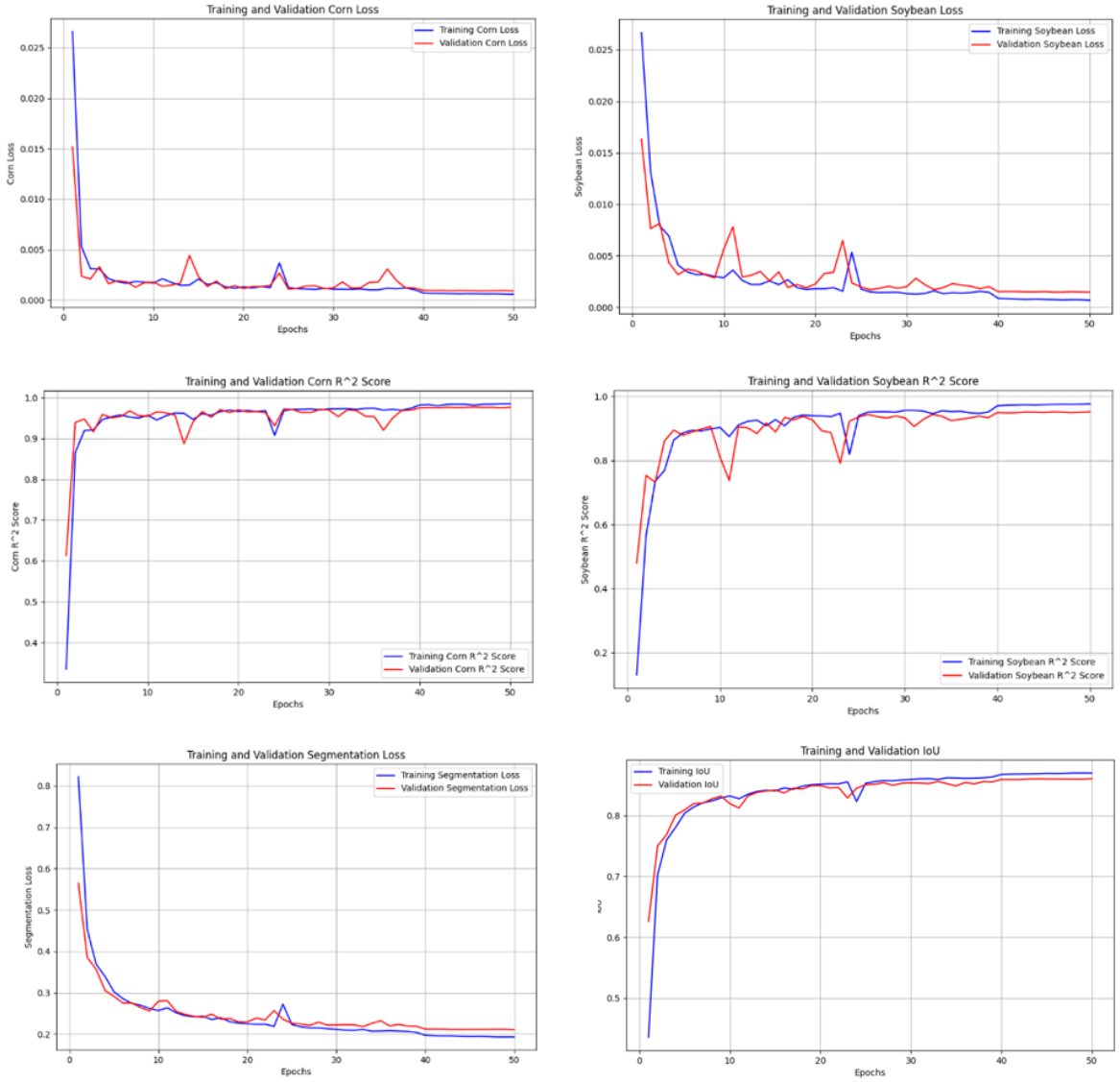


Figure 26: Training and validation learning curves of the multi-task learning Swin model (regression and segmentation)

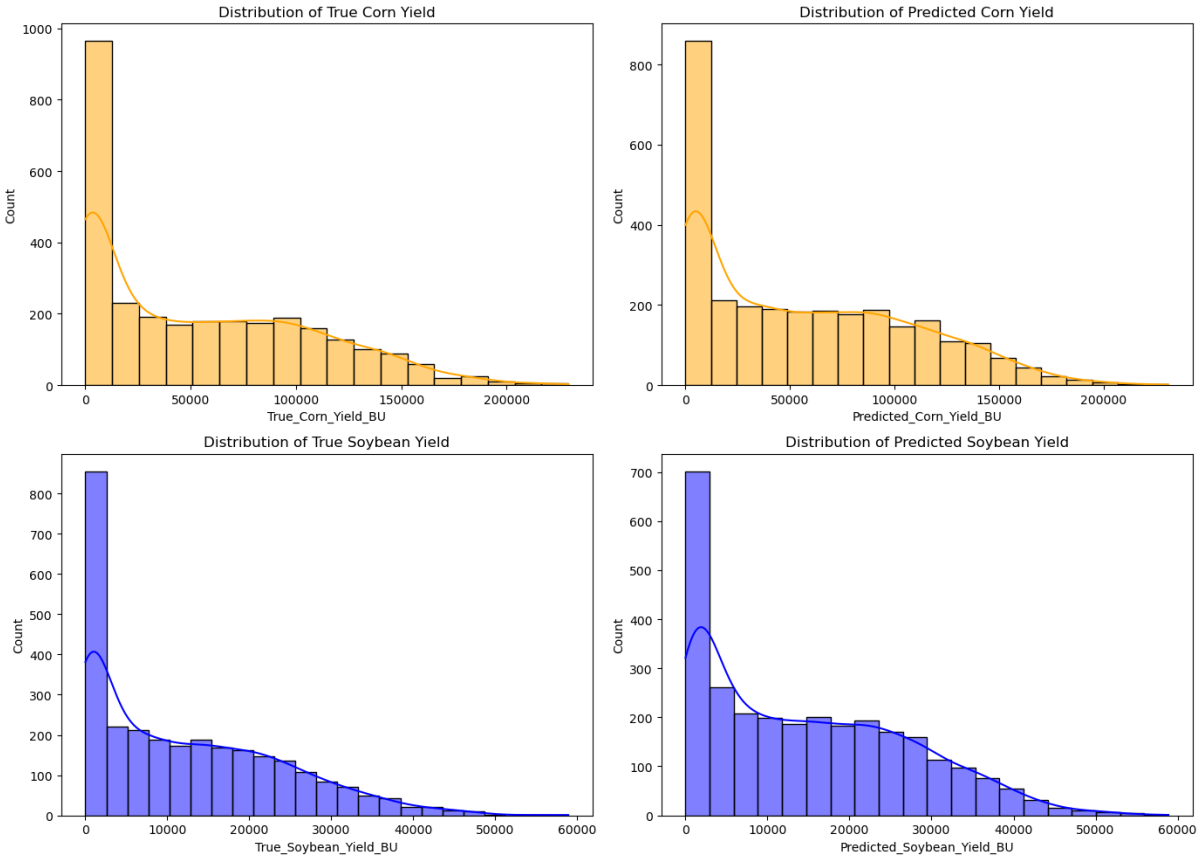


Figure 27: Distribution of corn and soybean yields of True values (left) and Predicted values (right), Minnesota-22 using the multi-task learning Swin model

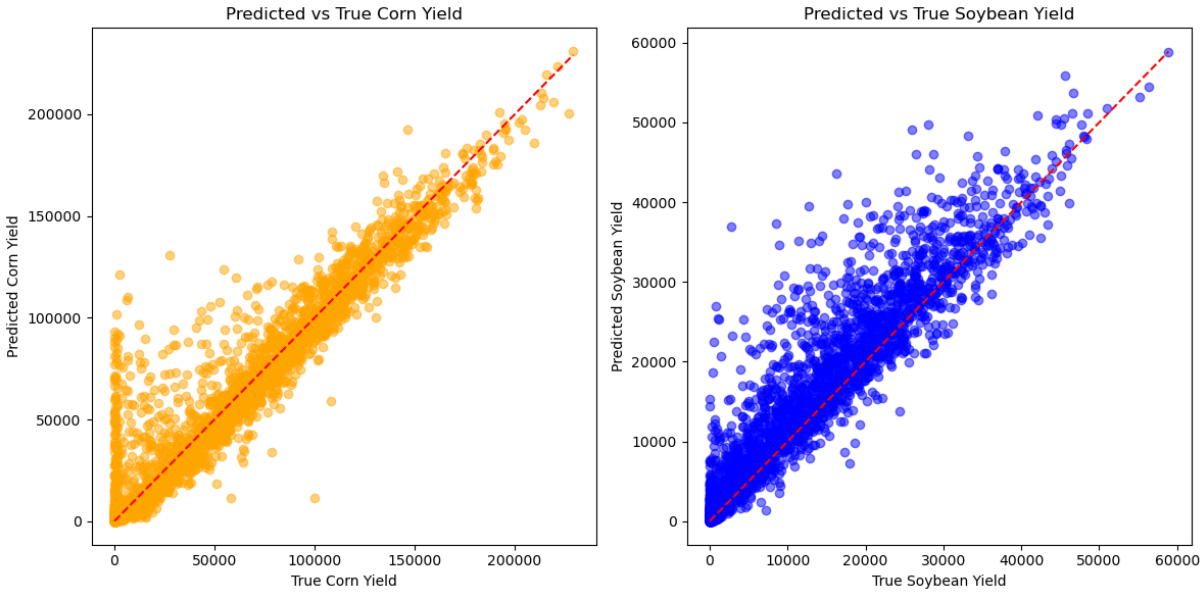


Figure 28: The 1:1 line of true and predicted yields of corn (left) and soybeans (right), Minnesota-22 using the multi-task learning Swin model

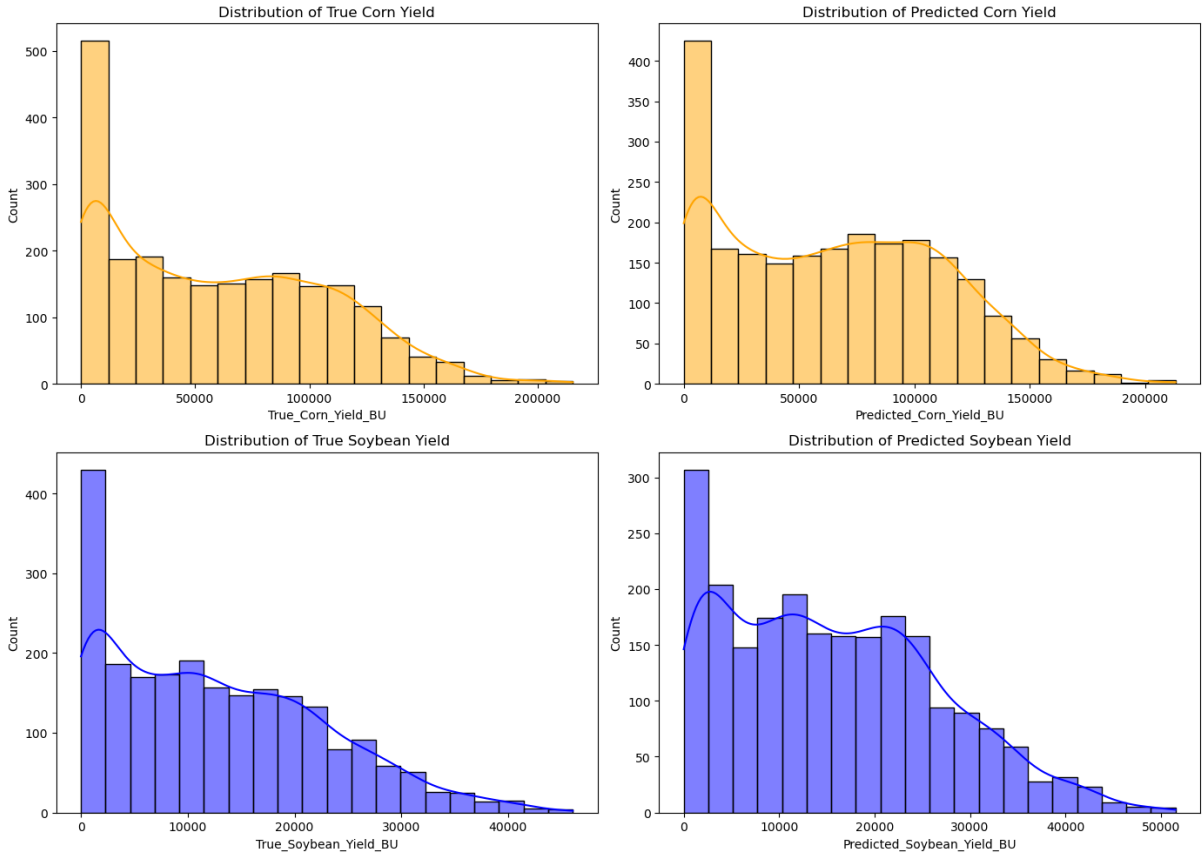


Figure 29: Distribution of corn and soybean yields of True values (left) and Predicted values (right), Minnesota-23 using the multi-task learning Swin model

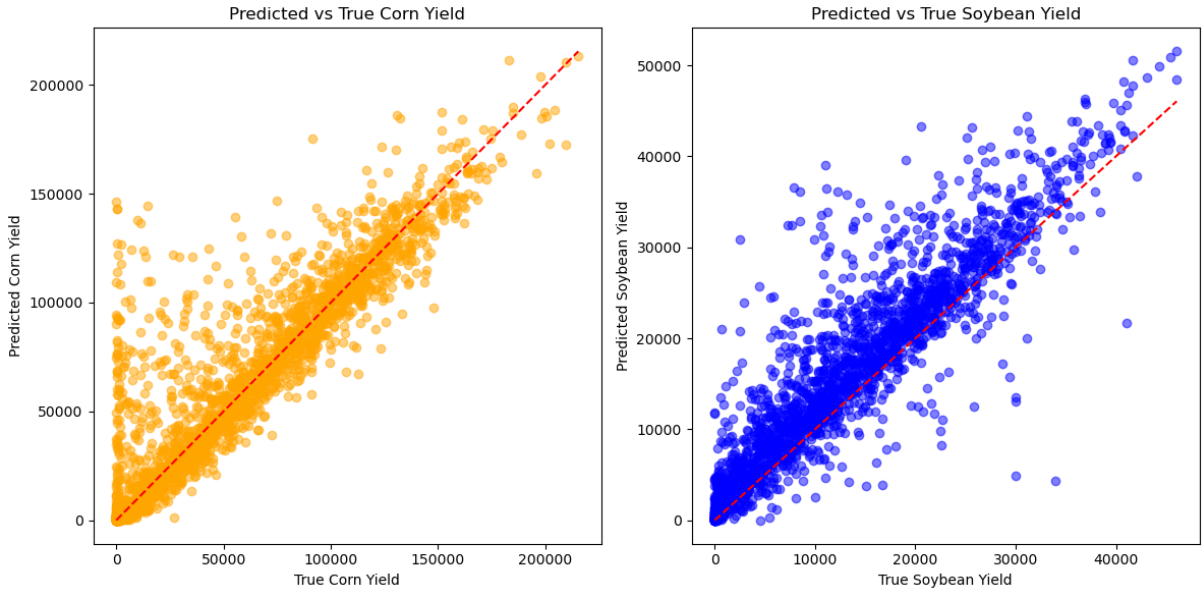


Figure 30: The 1:1 line of true and predicted yields of corn (left) and soybeans (right), Minnesota-23 using the multi-task learning Swin model

4.3. Comparisons

This section contains three distinct comparisons. Firstly, we compared the accuracy of the CNN model without CDL and the CNN model with CDL. Secondly, the regression evaluation metrics of all four developed models were compared. Thirdly, the testing results of the Swin and U-net models on Minnesota datasets from 2022 and 2023 were compared.

Table 11 contains the testing metrics of the four models based on Minnesota 2022 whereas Table 12 shows the testing metrics of only the multi-task learning models (U-net and Swin) on both the Minnesota 2022 and Minnesota 2023 test datasets. Figure 31 depicts the comparison between the CNN models with and without CDL based on the Minnesota 2022 test dataset in terms of RMSE and R^2. Figure 32 highlights the differences among the four models (CNN without CDL, CNN with CDL, U-net and Swin) on the Minnesota 2022 test dataset in terms of RMSE and R^2. Figure 33 demonstrates the difference between U-net and Swin on the Minnesota 2022 and Minnesota 2023 test datasets.

Looking at Table 11 and Figure 31, the RMSE significantly decreased for both corn and soybean, while the R^2 values significantly increased when using CDL as input to the CNN model. When the CDL was added, the RMSE reduced drastically from 24456.7 BU to 10468 BU and the R^2 significantly increased from 0.778 to 0.959 for corn. Furthermore, the RMSE reduced from 7724 BU to 3256 BU and the R^2 increased from 0.550 to 0.923 for the soybean.

As shown in Table 11 and Figure 32, the two multi-task learning models (U-net and Swin) achieved results that were remarkably close to those of the CNN with CDL and substantially higher than the CNN model without CDL. Additionally, the Swin model achieved slightly less accurate estimations for corn but slightly more accurate estimations for soybeans compared to the U-net model. Furthermore, the segmentation accuracy of the Swin model was higher than that of the U-net in terms of IOU.

As depicted in Table 12 and Figure 33, both regression and segmentation accuracies decreased when testing on the Minnesota 2023 test dataset. However, in both cases, the Swin model demonstrated higher segmentation accuracy and estimated soybean yield slightly more accurately than the U-net, while corn yield was estimated slightly less accurately than the U-net.

Table 11: Evaluation metrics of the two CNN models, U-net, and Swin in Minnesota 2022

	CNN (without CDL)		CNN (with CDL)		U-net		Swin	
	Corn	Soybean	Corn	Soybean	Corn	Soybean	Corn	Soybean
RMSE	24456.7 BU	7724 BU	10468 BU	3256 BU	14595 BU	5889 BU	15493 BU	5047 BU
R^2	0.778	0.550	0.959	0.923	0.921	0.738	0.911	0.808
IOU					0.7734		0.8001	
Notes	The average crop yield per patch of corn is 54064 BU and for soy is 12249 BU							

Table 12: Evaluation metrics of the multi-task learning U-net and Swin models in Minnesota 2022 and Minnesota 2023

		Minnesota_22 (MN22)	Minnesota_22 (MN22)	Minnesota_23 (MN23)	Minnesota_23 (MN23)
Metric	Crop	U-net	Swin	U-net	Swin
RMSE	Corn	14595	15493	19304	21842
RMSE	Soybean	5889	5047	5924	5374
R^2	Corn	0.921	0.911	0.836	0.79
R^2	Soybean	0.738	0.808	0.652	0.713
IOU		0.7734	0.8001	0.7194	0.7611

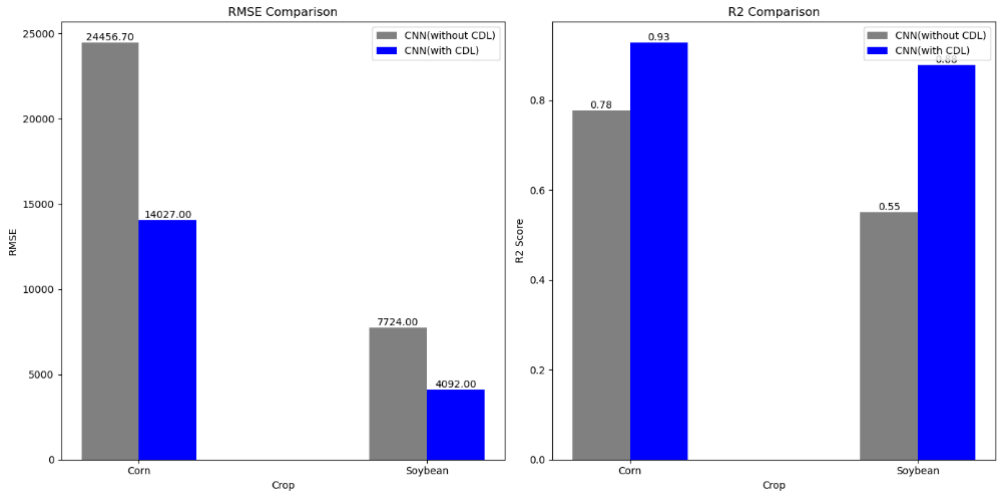


Figure 31: The comparison of CNN with and without CDL on Minnesota 2022

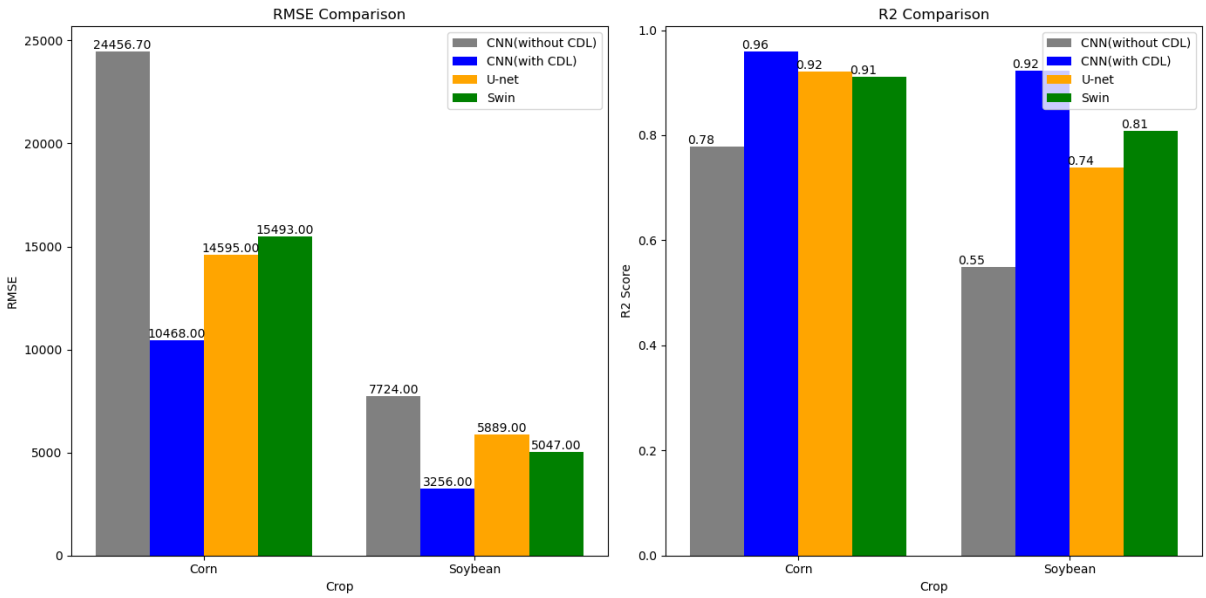


Figure 32: The comparison of the regression evaluation metric of all 4 models using RMSE (left) and R^2(right)

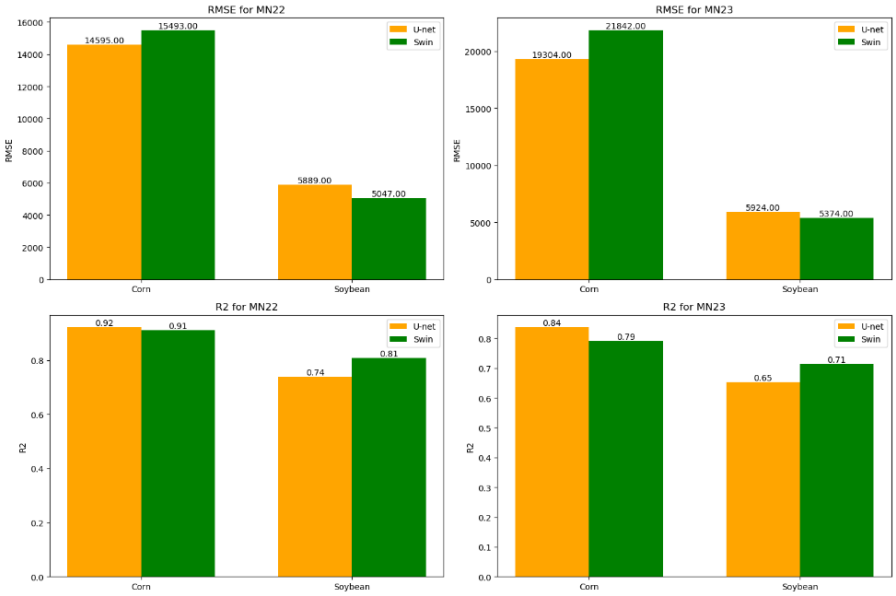


Figure 33: The comparison of U-net and Swin when testing on Minnesota 2022 and Minnesota 2023

4. DISCUSSION

This chapter discusses the outcomes of the developed models and highlights the key findings that help to answer the research objectives. The following sub-sections address four primary points. Firstly, we examined the impact of incorporating CDL as input in crop yield estimation models. Secondly, we explored the application of multi-task learning models for multi-crop yield estimation. Thirdly, we analysed the spatial and temporal generalizability of multi-task learning models for multi-crop yield estimation. Finally, we highlighted the common findings of the models.

4.1. Sub-objective 1

This section discusses the results affecting sub-objective 1 of assessing the impact of CDL on multi-crop yield estimation. A review of the results from sections 4.1.1 and 4.1.2, alongside the comparisons presented in Table 11 and Figure 31 in section 4.3, clearly indicates that incorporating CDL as input data substantially improves the multiple crops' estimation accuracy. This observation is consistent with Venugopal's (2023) conclusion on single-crop yield estimation (soybean) using CNNs, where saliency maps illustrated that CDL helps the models focus on relevant regions of the images. Specifically, CDL alone, when added to sentinel-2 images, significantly improved the RMSE for corn and soybean by 13,988 BU and 4,468 BU, respectively, and increased the R^2 values for corn and soybean by 0.181 and 0.373, respectively.

4.2. Sub-objective 2

This section discusses the relevant results to sub-objective 2 of assessing the applicability of multi-task learning models on multi-crop yield estimation. As detailed in sections 3.4.2 and 3.4.3, two multi-task learning models were successfully developed and implemented for simultaneous crop type identification and yield estimation. The first model utilized U-net as its backbone, while the second employed Swin. The results presented in Table 11 and Figure 32 indicate that both models achieved regression results comparable to the CNN model with CDL and significantly outperformed the CNN model without CDL. Regarding segmentation accuracy, Swin demonstrated superior performance with an IOU of 0.8001, compared to U-net's IOU of 0.7734. Furthermore, Swin exhibited higher accuracy in estimating soybean yields compared to U-net, while achieving slightly lower accuracy in estimating corn yields.

4.3. Sub-objective 3

This section discusses the results related to sub-objective 3 of assessing spatial and temporal generalizability of multi-task learning models on multi-crop yield estimation. By comparing the results of the multi-task models with the CNN model incorporating CDL on the Minnesota 2022 dataset, these models showed strong spatial generalizability. Additionally, they maintained relatively good temporal performance when tested on the Minnesota 2023 dataset. However, there was a slight decline in both regression and segmentation accuracies for Minnesota 2023. As shown in Table 12 and Figure 33, the RMSE for corn and soybean increased by 4,709 BU and 35 BU, respectively, for U-net, and by 6,349 BU and 1,327 BU, respectively, for Swin. Furthermore, the R^2 values for corn and soybean decreased by 0.085 and 0.086 for U-net, and by 0.121 and 0.095 for Swin. This indicates that U-net demonstrated better temporal generalizability in regression tasks compared to Swin with the given data size. Conversely, Swin achieved higher segmentation accuracies for both datasets (Minnesota 2022 and 2023). The drop with Minnesota 2023 in IOU for U-net was 0.054, whereas for Swin it was 0.039. This suggests that Swin demonstrated better temporal generalizability in segmentation tasks than U-net with the data size used.

4.4. Common discussion points

This section discusses common findings on all the developed models. The results of all the developed models indicate that farms with extremely low crop yield or near-zero yield values can confuse the models, resulting in either negative values or excessively high values. This phenomenon is illustrated in Figures 17, 20, 23, 25, 28, and 30, which show the 1:1 line between predicted and target yield values of corn and soybean. These observations suggest a need for further research to address low-production farms or to develop specific models for cases with below-average crop yields. Additionally, transformer-based models, unlike CNNs, require extensive training data due to their lack of inherent inductive biases (i.e., assuming preliminary knowledge about the image data such as weight sharing and translation invariance) (Khan et al., 2022). Therefore, it is anticipated that Swin would achieve higher accuracy with a larger dataset. Furthermore, due to time constraints, the models were trained for only 50 epochs. Increasing the number of epochs could potentially improve accuracy. All the models' results were obtained by training on data from a single year (2022) and using the average reflectance values of Sentinel-2 images from July and August. Expanding the dataset to include multiple years and different periods within the growing season is expected to enhance the temporal generalizability of the models. Finally, it is observed that the accuracy of the estimated corn yield in all models is higher than in soybean, this could be because the average crop yield per input image for corn is higher than that of soybean in all the datasets as shown in Table 4. Furthermore, the mid-season of corn spans from June to August, while for soybean, it spans from July to August as depicted in Figure 10. However, the reflectance values in this study were averaged in July and August. This could contribute to the low accuracy of soybeans compared to corn.

Furthermore, Although the crop yield data provided by the USDA are at the county level, we downsampled this data to the pixel level and then aggregated them to the image level. This method may present a limitation by estimating crop yield at a finer level than the original data. However, one could argue that if the USDA data are provided as bushels per acre (BU/Acre) per county, then this value can be applied to any acre within that county to calculate the total crop yield for a specific area. This approach allowed us to obtain the total crop yield per input image. Furthermore, it significantly increased the training data, thus enabling the training of multiple DL models. Importantly, the DL models trained on the downsampled data still produced satisfactory results.

5. CONCLUSIONS

This chapter is divided into three sections. The first section reviews the research objectives and questions, highlighting that all sub-objectives were achieved, and the research questions were answered. The second section offers future recommendations to enhance our developed models. The third section provides an overall conclusion derived from our research.

5.1. Research Objectives achieved and Research questions

The main objective was achieved through multi-task learning models, which identified crop types and estimated their yields. Table 13 illustrates the sub-objectives and the answers to each of their questions.

Table 13: Research sub-objectives achieved and questions answered

	Sub-objectives	Research questions and answers	
1	**Assess the CDL effect on multi-crop yield estimation using a base CNN model.**	RQ 1.1	**How effective in terms of accuracy is adding the CDL as a factor for estimating multi-crop yield?**
		Answer 1.1	Incorporating CDL into the CNN model reduced RMSE by 10,429 BU for corn and 3,632 BU for soybean. Additionally, the R^2 value for corn increased by 0.15, and for soybean, it increased by 0.329.
2	**Develop two multi-task learning models for crop type identification and crop yield estimation. The first is UNET-based. The second is SWIN-based.**	RQ 2.1	**Is it feasible to use segmentation models such as U-net and Swin as a backbone for both segmentation and regression**?
		Answer 2.1	Yes, two multi-task learning models based on the segmentation models of U-net and Swin were successfully developed for crop type identification and yield estimation.
		RQ 2.2	**Can multi-task learning models achieve multi-crop yield accuracies comparable to CNNs with CDL as input?**
		Answer 2.2	Yes, the two developed multi-task learning models achieved reliable results compared to the CNN with CDL as input as detailed in Table (11) and Figure 32.
3	**Assess the performance of multi-task learning models**	RQ 3.1	**How accurately can multi-task learning models generalize spatially and temporally?**

	on unseen regions and unseen years.	Answer 3.1	Both developed multi-task learning models demonstrated strong spatial and temporal performance, with spatial results outperforming temporal ones. This discrepancy may be attributed to the models being trained on data from a single year.

5.2. Future Recommendations

To identify the limitations of this research and highlight possible future work, we categorized the future recommendations into three sections: Data-related, Acquisition-date-related, and models-related aspects. Table 14 shows each category and lists the points of future work per each. It also gives a brief explanation of each point.

Table 14: Future recommendations

Category	Point of work	Explanation
Data-Related	Selected bands	we used 8 bands of sentinel-2. However, experimenting with different bands that have more influence on vegetation could potentially enhance accuracy.
	Different sensors	different sensors with different spectral bands, spatial resolution, and temporal resolution could be chosen.
	Additional data	adding extra data that are important in crop yield estimations such as weather and soil type data could enhance the model's generalizability.
	Larger data size	models such as transformers are data hungry. Therefore, increasing the data size could potentially enhance the accuracy of the models.
	Crop yield average in different regions	Further exploration of the distribution of crop yield values and crop type classes and their effect on the generalizability of the model is important to use the developed models for far regions and different countries.
Acquisition date-related	Mid-season data	It is recommended to experiment with different acquisition dates in mid-season rather than averaging the whole mid-season reflectance values.
	Temporal data	It is recommended to train on data from different years.
Models-related	Activation functions	It is recommended to experiment with different activation functions for segmentation and regression.
	Combining activation functions	It is recommended to experiment with different combined loss functions in multi-task learning models instead of only adding all the losses.

	Hyperparameter tuning	It is recommended to perform hyperparameter tuning to enhance the models' results.

5.3. Overall conclusion

Crop yield estimation is crucial for commodity management and ensuring food security. While most research in this field has focused on estimating single-crop yields, this does not accurately reflect the reality of multiple crops being grown in the same area This MSc thesis began by verifying previous findings that indicated that using CDL as input enhances the accuracy of crop yield estimation, but we extended this verification to multiple crops. However, creating the CDL is time-consuming, and it is not available in all countries. Consequently, we experimented with multi-task models capable of concurrently identifying crop types and estimating yields. This approach leveraged transfer learning among tasks to improve the accuracy of both segmentation and regression tasks. Two models for segmentation (U-net and Swin) were employed as backbones for feature extraction. Each model was equipped with two heads: one for segmentation to identify crop types, and one for regression to estimate crop yields. Both models demonstrated accuracies comparable to the CNN with CDL, suggesting that multi-task learning models for crop type identification and yield estimation can serve as effective tools for multi-crop yield estimation, eliminating the need for CDL. Although these models showed good spatial and temporal performance when tested on Minnesota data from 2022 and 2023, there remains room for improvement. These potential enhancements were categorized into three areas, as detailed in section 5.2, Table 14. In conclusion, the main contribution of this research is the development of multi-task learning models. These models proved their applicability as actionable models that can be directly used for multi-crop yield estimation while achieving good results. Furthermore, they addressed the challenge posed by the unavailability of the CDL as input. Importantly, this is done using one model which reduces time, computational power, and learned models' parameters compared to developing single models for each crop.

6. LIST OF REFERENCES

Azad, R., Heidary, M., Yilmaz, K., Hüttemann, M., Karimijafarbigloo, S., Wu, Y., Schmeink, A., & Merhof, D. (2023). *Loss Functions in the Era of Semantic Segmentation: A Survey and Outlook*. http://arxiv.org/abs/2312.05391

Bi, L., Wally, O., Hu, G., Tenuta, A. U., Kandel, Y. R., & Mueller, D. S. (2023). A transformer-based approach for early prediction of soybean yield using time-series images. *Frontiers in Plant Science*, *14*, 1173036. https://doi.org/10.3389/FPLS.2023.1173036/BIBTEX

Caruana, R. (1997). Multitask Learning. *Machine Learning*, *28*(1), 41–75. https://doi.org/10.1023/A:1007379606734/METRICS

Chang, C.-H. ;, Lin, J. ;, Chang, J.-W. ;, Huang, Y.-S. ;, Lai, M.-H. ;, Chang, Y.-J., Chang, C.-H., Lin, J., Chang, J.-W., Huang, Y.-S., Lai, M.-H., & Chang, Y.-J. (2024). Hybrid Deep Neural Networks with Multi-Tasking for Rice Yield Prediction Using Remote Sensing Data. *Agriculture 2024, Vol. 14, Page 513*, *14*(4), 513. https://doi.org/10.3390/AGRICULTURE14040513

Cipolla, R., Gal, Y., & Kendall, A. (2017). Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 7482–7491. https://doi.org/10.1109/CVPR.2018.00781

Crawshaw, M. (2020). *Multi-Task Learning with Deep Neural Networks: A Survey*. https://arxiv.org/abs/2009.09796v1

Desloires, J., Ienco, D., & Botrel, A. (2023). Out-of-year corn yield prediction at field-scale using Sentinel-2 satellite imagery and machine learning methods. *Computers and Electronics in Agriculture*, *209*, 107807. https://doi.org/10.1016/J.COMPAG.2023.107807

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*, 4171–4186. https://arxiv.org/abs/1810.04805v2

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. http://arxiv.org/abs/2010.11929

Ebert, N., Mangat, P., & Wasenmuller, O. (2022). Multitask Network for Joint Object Detection, Semantic Segmentation and Human Pose Estimation in Vehicle Occupancy Monitoring. *IEEE Intelligent Vehicles Symposium, Proceedings*, *2022-June*, 637–643. https://doi.org/10.1109/IV51971.2022.9827088

Htun, A. M., Shamsuzzoha, M., & Ahamed, T. (2023). Rice yield prediction model using normalized vegetation and water indices from Sentinel-2A satellite imagery datasets. *Asia-Pacific Journal of Regional Science*, *7*(2), 491–519. https://doi.org/10.1007/S41685-023-00299-2

Ilyas, Q. M., Ahmad, M., & Mehmood, A. (2023). Automated Estimation of Crop Yield Using Artificial Intelligence and Remote Sensing Technologies. *Bioengineering*, *10*(2). https://doi.org/10.3390/BIOENGINEERING10020125

Jadon, A., Patil, A., & Jadon, S. (2022). *A Comprehensive Survey of Regression Based Loss Functions for Time Series Forecasting*. http://arxiv.org/abs/2211.02989

Jhajharia, K., & Mathur, P. (2023). Prediction of crop yield using satellite vegetation indices combined with machine learning approaches. *Advances in Space Research*, *72*(9), 3998–4007. https://doi.org/10.1016/j.asr.2023.07.006

Johnson, M. D., Hsieh, W. W., Cannon, A. J., Davidson, A., & Bédard, F. (2016). Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agricultural and Forest Meteorology*, *218–219*, 74–84. https://doi.org/10.1016/J.AGRFORMET.2015.11.003

Joshi, A., Pradhan, B., Chakraborty, S., & Behera, M. D. (2023). Winter wheat yield prediction in the conterminous United States using solar-induced chlorophyll fluorescence data and XGBoost and random forest algorithm. *Ecological Informatics*, *77*, 102194. https://doi.org/10.1016/J.ECOINF.2023.102194

Kaplan, G., & Avdan, U. (2017). Object-based water body extraction model using Sentinel-2 satellite imagery. *European Journal of Remote Sensing*, *50*(1), 137–143. https://doi.org/10.1080/22797254.2017.1297540

Khaki, S., Pham, H., & Wang, L. (2021). Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning. *Scientific Reports*, *11*(1). https://doi.org/10.1038/S41598-021-89779-Z

Khan, S., Naseer, M., Khan, S., Naseer, M., City, M., Dhabi, A., Zamir, S. W., Shah, M., Hayat, M., Waqas Zamir, S., Shahbaz Khan, F., & Shah, M. (2022). Transformers in Vision: A Survey; Transformers in Vision: A Survey. *ACM Computing Surveys*, *54*(10s). https://doi.org/10.1145/3505244

Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. http://arxiv.org/abs/1412.6980

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine Learning in Agriculture: A Review. *Sensors 2018, Vol. 18, Page 2674*, *18*(8), 2674. https://doi.org/10.3390/S18082674

Lin, F., Crawford, S., Guillot, K., Zhang, Y., Chen, Y., Yuan, X., Chen, L., Williams, S., Minvielle, R., Xiao, X., Gholson, D., Ashwell, N., Setiyono, T., Tubana, B., Peng, L., Bayoumi, M., & Tzeng, N.-F. (2023). MMST-ViT: Climate Change-aware Crop Yield Prediction via Multi-Modal Spatial-Temporal Vision Transformer. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 5751–5761. https://doi.org/10.1109/ICCV51070.2023.00531

Liu, X., & Wang, H. (2019). AdvNet: Multi-Task Fusion of Object Detection and Semantic Segmentation. *Proceedings - 2019 Chinese Automation Congress, CAC 2019*, 3359–3362. https://doi.org/10.1109/CAC48633.2019.8997277

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986

Marshall, M., Belgiu, M., Boschetti, M., Pepe, M., Stein, A., & Nelson, A. (2022). Field-level crop yield estimation with PRISMA and Sentinel-2. *ISPRS Journal of Photogrammetry and Remote Sensing*, *187*, 191–210. https://doi.org/10.1016/J.ISPRSJPRS.2022.03.008

Mohammadi, S. (2024). *Crop type mapping from satellite image time series using deep learning* [University of Twente]. https://doi.org/10.3990/1.9789036560993

Oikonomidis, A., Catal, C., & Kassahun, A. (2023). Deep learning for crop yield prediction: a systematic literature review. *New Zealand Journal of Crop and Horticultural Science*, *51*(1), 1–26. https://doi.org/10.1080/01140671.2022.2032213

Overview — MMSegmentation 1.2.2 documentation. (2024, June 24). https://mmsegmentation.readthedocs.io/en/latest/overview.html

ReduceLROnPlateau — PyTorch 2.3 documentation. (2024, June 14). https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html#reducelronplateau

Rizwan I Haque, I., & Neubert, J. (2020). Deep learning approaches to biomedical image segmentation. *Informatics in Medicine Unlocked*, *18*, 100297. https://doi.org/10.1016/J.IMU.2020.100297

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. http://arxiv.org/abs/1505.04597

Sagan, V., Maimaitijiang, M., Bhadra, S., Maimaitiyiming, M., Brown, D. R., Sidike, P., & Fritschi, F. B. (2021). Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *174*, 265–281. https://doi.org/10.1016/J.ISPRSJPRS.2021.02.008

Srivastava, A. K., Safaei, N., Khaki, S., Lopez, G., Zeng, W., Ewert, F., Gaiser, T., & Rahimi, J. (2022). Winter wheat yield prediction using convolutional neural networks from environmental and phenological data. *Scientific Reports 2022 12:1*, *12*(1), 1–14. https://doi.org/10.1038/s41598-022-06249-w

Sun, J., Di, L., Sun, Z., Shen, Y., & Lai, Z. (2019). County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model. *Sensors 2019, Vol. 19, Page 4363*, *19*(20), 4363. https://doi.org/10.3390/S19204363

Sun, Z., Li, Q., Jin, S., Song, Y., Xu, S., Wang, X., Cai, J., Zhou, Q., Ge, Y., Zhang, R., Zang, J., & Jiang, D. (2022a). Simultaneous Prediction of Wheat Yield and Grain Protein Content Using Multitask Deep Learning from Time-Series Proximal Sensing. *Plant Phenomics*, *2022*. https://doi.org/10.34133/2022/9757948

Sun, Z., Li, Q., Jin, S., Song, Y., Xu, S., Wang, X., Cai, J., Zhou, Q., Ge, Y., Zhang, R., Zang, J., & Jiang, D. (2022b). Simultaneous Prediction of Wheat Yield and Grain Protein Content Using Multitask Deep Learning from Time-Series Proximal Sensing. *Plant Phenomics*, *2022*. https://doi.org/10.34133/2022/9757948

Tatachar, A. V. (2021). Comparative Assessment of Regression Models Based On Model Evaluation Metrics. *International Research Journal of Engineering and Technology*. www.irjet.net

United States - Crop Calendar. (2024, June 16). https://ipad.fas.usda.gov/rssiws/al/crop_calendar/us.aspx

US Soybean Production by State: Ranking the Top 11. (2024, June 16). https://www.cropprophet.com/soybean-production-by-state-top-11/

USDA - National Agricultural Statistics Service - Research and Science - Cropland Data Layers. (2024, June 16). https://www.nass.usda.gov/Research_and_Science/Cropland/sarsfaqs2.php

van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, *177*, 105709. https://doi.org/10.1016/J.COMPAG.2020.105709

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*, *2017-December*, 5999–6009. https://arxiv.org/abs/1706.03762v7

Venugopal, A. (2023). *A regression-based Convolutional Neural Network for yield estimation of soybean*. https://essay.utwente.nl/96269/

Wang, N., Ma, Z., Huo, P., Liu, X., He, Z., & Lu, K. (2023). 3D Convolutional Neural Network with Dimension Reduction and Metric Learning for Crop Yield Prediction Based on Remote Sensing Data. *Applied Sciences 2023, Vol. 13, Page 13305*, *13*(24), 13305. https://doi.org/10.3390/APP132413305

Zhang, H., Zhang, Y., Liu, K., Lan, S., Gao, T., & Li, M. (2023). Winter wheat yield prediction using integrated Landsat 8 and Sentinel-2 vegetation index time-series data and machine learning algorithms. *Computers and Electronics in Agriculture*, *213*, 108250. https://doi.org/10.1016/J.COMPAG.2023.108250

Zhang, L., Zhang, Z., Luo, Y., Cao, J., Xie, R., & Li, S. (2021). Integrating satellite-derived climatic and vegetation indices to predict smallholder maize yield using deep learning. *Agricultural and Forest Meteorology*, *311*, 108666. https://doi.org/10.1016/J.AGRFORMET.2021.108666