Spectral Satellite Image Analysis for Crop Classification

Kevin Cheung
1909190

September 2023

Swansea University
Prifysgol Abertawe

Project Dissertation submitted to Swansea University in the Partial Fulfilment for the Degree of Master of Science

Department of Computer Science

SPECTRAL SATELLITE IMAGE ANALYSIS FOR CROP CLASSIFICATION

ABSTRACT

Assessing crop yields and ensuring food security is crucial as the global population continues to grow. However, accurate estimation remains a challenge due to limited available data. To address this issue, agricultural monitoring techniques have been developed, leveraging remote sensing data from satellites and machine learning.

Traditional crop classification models often focus on either spatial or temporal data. This research introduces a deep learning model designed to harness information from both spatial and temporal domains. Throughout the project, investigations will be conducted to determine optimal image preprocessing techniques, employ data augmentation methods, and identify the most influential spectral bands and vegetation indices for the model's performance.

Moreover, this study explores the impact of varying time series lengths in the dataset. Conventional crop classification techniques rely on year-long datasets for crop type classification, which may not align with estimating potential crop yield as the harvest would have already occurred. This research aims to analyse how prediction accuracy varies at different stages of the crop growth cycle, addressing a critical aspect of agricultural monitoring.

DECLARATIONS

DECLARATION
This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

STATEMENT 1
This dissertation is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by giving explicit references. A bibliography is appended.

STATEMENT 2
I hereby give consent for my dissertation, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organizations.

Signed: Kevin Cheung
Date: 1/10/23

Contents

List of Figures

1 INTRODUCTION

Pressure on agriculture grows with the increasing global population along with rising rates of consumption. To meet this demand, both small and large scale farms must increase the productivity in order to ensure food security for the world. However, the quantities of crops being grown have been degrading land, water, biodiversity and climate on a global scale[1]. Causes of these issues could be caused by poor agricultural practices, such as intensive irrigation that causes water to infiltrate the upper levels of soil and leach the nutrients deeper into the ground while also drowning any organisms living there[2]. Knowledge of the field extent, crop type and distribution across farmland is important for managing the agricultural expansion in a sustainable manner[3] but this process requires regular agricultural monitoring. Such monitoring can be used to estimate the crop yield, map the land use of the crops or calculate the green house emissions emitted from the farmland, which can also be used to manage the subsidies paid to the farmers[4].

The Common Agricultural Policy (CAP) in the EU requires farmers to farmers to submit an application detailing the precise details of the agricultural parcels along with the crop type for the farmers to receive financial support. These applications must be submitted accurately and early in the year otherwise, the member state must return part of the subsidy received by the EU[5]. This data is verified using a team of photo interpreters, which would be time consuming, so there is a need to utilise an automated crop classification system. Furthermore, farmers can use the crop detection images to monitor their farmland. A visual inspection of a farm can provide a limited view of the farm, but information such as the crop stress or moisture levels of inaccessible areas of a crop can also be assessed. These tools can also be used on smaller subsistence farms around the world. There are 1.5 billion people relying on subsistence farms around the world who have limited access to knowledge, assets, credit, markets, and risk management that can come from larger-scale agricultural enterprises [6], so being able to automate the process of crop classification will allow access to data on the food security in regions where limited data is collected. Furthermore, if smaller businesses do not have access to these tools, the inequalities between small and commercial farms will be exacerbated and therefore increases the difficulty of food security.

1.1 Project Description

Crop classification is the process of identifying and categorizing different types of crops or vegetation in a specific area based on remote sensing data obtained from sources like satellite imagery or aerial photography. Remote sensing involves the use of sensors to capture information about the Earth's

surface by analyzing reflected electromagnetic radiation, resulting in the generation of images [7]. It has diverse applications in fields such as environmental monitoring, weather forecasting, urban planning, and, in the context of this study, agricultural and crop management. Through remote sensing data, farmers gain insights into various aspects of their farm areas, including moisture levels, disease outbreaks, and damage assessment [8].

Remote sensing equipment is capable of collecting data across the entire electromagnetic spectrum. In this study, Sentinel-2 remote sensing data was utilised, encompassing information from 13 spectral bands [9]. Sentinel-2 employs a passive sensor to detect electromagnetic waves originating from the sun and reflected by the Earth's surface. Notably, the waves absorbed and reflected by vegetation carry unique signatures for each plant species [10], enabling the creation of spectral fingerprints that characterise unknown vegetation types.

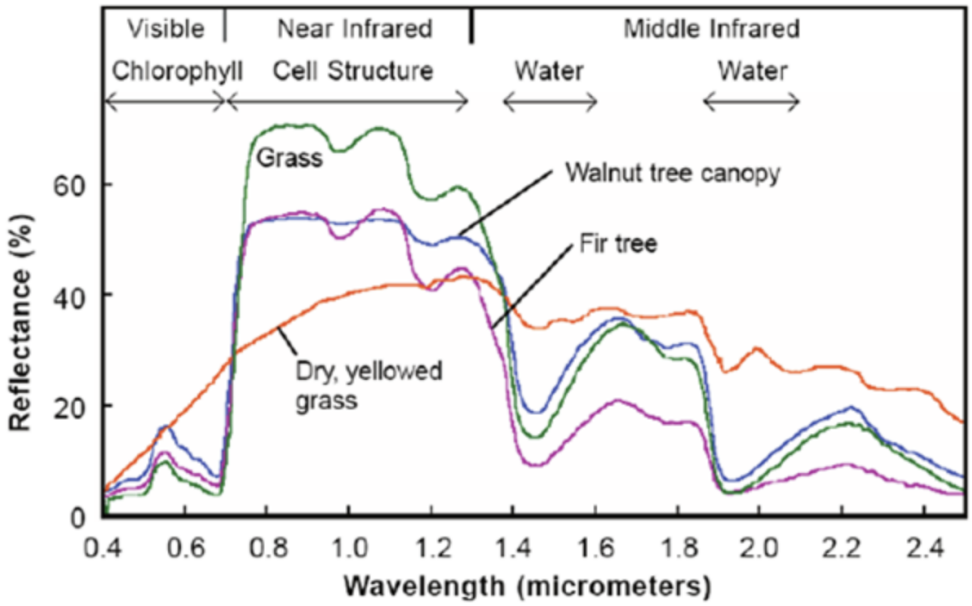


Fig. 1. Spectral signature of different types of vegetation

Figure 1 [11] presents a collection of spectral fingerprints for various plant types, spanning wavelengths from visible light to infrared. This figure also highlights the specific wavelength ranges influenced by factors like chlorophyll, cell structure, and water content. These details provide valuable guidance on selecting the most suitable spectral bands for classification purposes. Considering factors such as cloud cover and moisture variations in vegetation, the optimal spectral region for crop classification lies within the visible and near-infrared range. The dataset utilised in this project comprises remote sensing data sourced from Sentinel-2 [12], encompassing the region of Brandenburg, Germany. This dataset encompasses an array of labeled crop regions, exceeding 2500

in number, spanning nine distinct crop types and spanning the timeframe between 2018 to 2019. Leveraging this dataset, the project aims to employ deep learning methodologies to facilitate crop classification via computer vision techniques. Computer vision, a subfield of artificial intelligence, is devoted to empowering computers with the capability to interpret and comprehend visual information. Within the realm of deep learning, there exist three fundamental categories: supervised learning, unsupervised learning, and semi-supervised learning. For the purposes of this project, a supervised learning approach will be adopted, capitalising on the labeled data provided within the dataset to train and refine the model for accurate crop classification. Using deep learning, this paper will utilise mutliclass and binary classification techniques for crop classification.

1.2 Research Aims

The primary objective of this project is to develop a robust deep learning model for crop classification, utilising the provided dataset. This study is centered on constructing a model with the capacity to analyse data in both spatial and temporal dimensions, employing supervised deep learning techniques. Additionally, the model will be subjected to training on time series data of varying lengths to assess its effectiveness in accurately classifying crops throughout their growth cycles.

1.3 Research Objectives

(1) Conduct a comprehensive review of the existing literature to gain insights into the latest advancements and achievements in the application of deep learning for agricultural monitoring.
(2) Develop an initial project proposal outlining your approach and strategies for building a crop classification model.
(3) Learn how to create deep learning models using Pytorch and Python.
(4) Preprocess the images in the dataset to apply image augmentation and transformations
(5) Implement the proposed crop classification for multiclass classification
(6) Evaluate and iterate on the model to achieve optimal results
(7) Implement the proposed crop classification for binary classification
(8) Produce results through experimentation on varying lengths of time series data
(9) Evaluate the results of the experiment

2 BACKGROUND

2.1 Remote Sensing

Remote sensing is a technique using sensors on airborne or satellites to collect information over an area or object by providing data about the objects at the Earth surface or atmosphere based on the radiation reflected or emitted from objects or areas in multiscale and multitemporal approaches [13]. Remote sensing can be categorised into two groups, active remote sensing and passive remote sensing. An active sensing system uses its own source of energy, which is directed at the area or object, to measure the return signal. An example of this would be RADAR or LIDAR. An example of satellite using an active remote sensor would be Sentinel-1 [14] which uses a Synthetic Aperture Radar (SAR) [15]. On the other hand, passive remote sensing utilises sensors that detect reflected or emitted electromagnetic radiation from external sources such as the sun. Most passive systems used by remote sensing applications operate in the visible, infrared, thermal infrared, and microwave portions of the electromagnetic spectrum [16]. However most of the wavelengths that utilise passive sensors detect cannot penetrate dense cloud cover thus limiting its effectiveness in areas like the tropics where dense cloud cover is frequent.

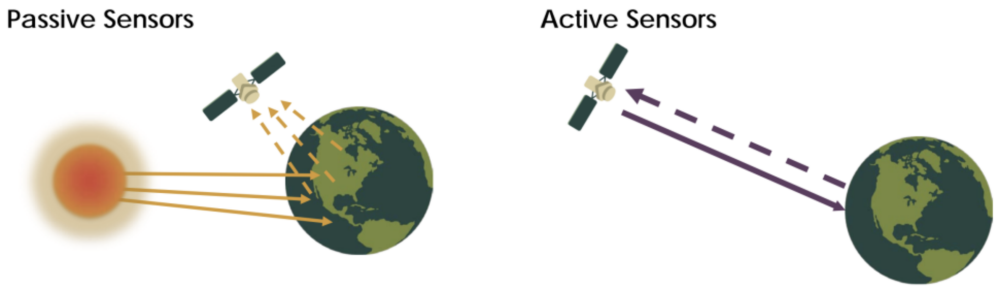


Fig. 2. Passive and Active remote sensing

Electromagnetic radiation manifests as waves that propagate through the atmosphere and the vacuum of space, encompassing a broad spectrum of wavelengths and frequencies. These characteristics define the diverse forms of electromagnetic radiation. At one end of the spectrum, we encounter shorter wavelengths with high frequencies, including ultraviolet, x-rays, and gamma rays. Conversely, longer wavelengths characterise the other end, housing radio waves, microwaves, and infrared radiation. In the middle of this spectrum resides visible light, perceptible to the human eye without the need for special instrumentation.

While visible light is within the realm of human perception, the detection of other electromagnetic wave types necessitates specialised instruments, such as remote sensors. These sensors are engineered

to capture data across the entire electromagnetic spectrum. The advantage of having access to this extensive range of wavelengths lies in the varying interactions of these waves with objects on the Earth's surface and atmospheric conditions. Some examples of this can be seen in Figure 3.

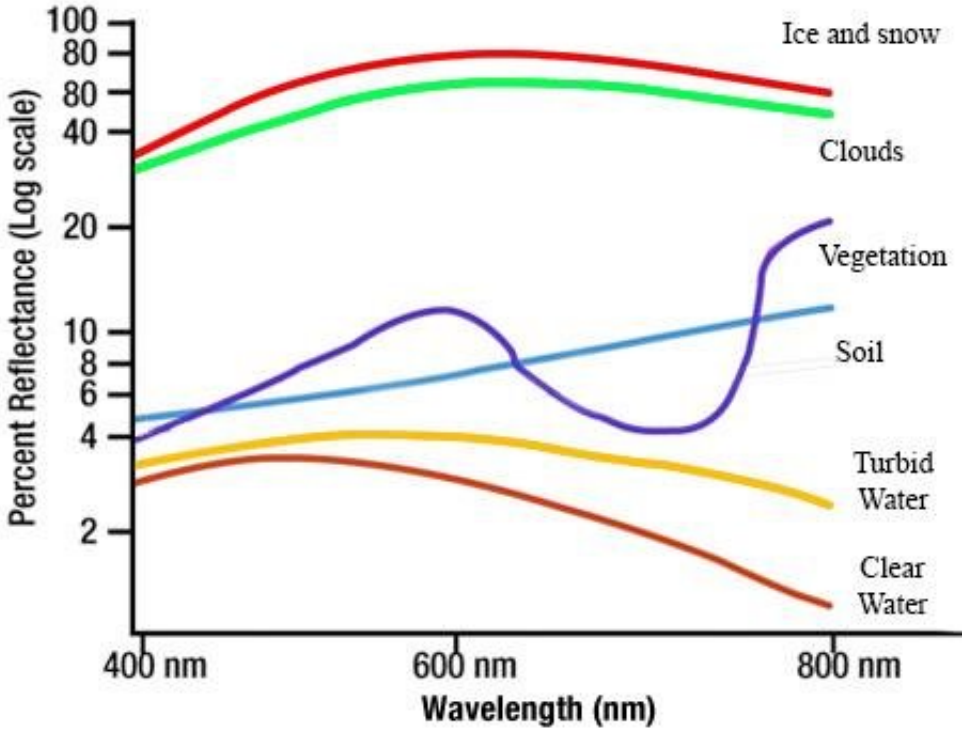


Fig. 3. Reflectance of light of spectral signatures of Earth features

For instance, certain wavelengths are more prone to reflectance off surfaces, while others are readily absorbed. This distinct behavior across the spectrum allows for a comprehensive understanding of our surroundings. Notably, atmospheric factors, like water vapor, can obstruct the transmission of visible wavelengths. However, microwaves, with their higher frequencies and lower wavelength, can effectively penetrate through cloud cover [17], enabling remote sensing instruments to collect valuable data even under adverse weather conditions. This adaptability and versatility in wavelength selection empower remote sensing technologies to provide crucial insights into our environment and facilitate a wide range of applications, from weather forecasting to agriculture and environmental monitoring.

In this research paper, remote sensing data is sourced from two key satellites, Sentinel-2A and Sentinel-2B, which are integral components of the Copernicus Programme [18]. These two satellites are strategically positioned in Earth's orbit, phased at 180 degrees apart from each other. This orbital

arrangement ensures that, at any given moment, one of the satellites will be located on the opposite side of the Earth compared to the other.

Each of these Sentinel satellites has a revisit time of 10 days[12], signifying the frequency with which they pass over and collect data from specific regions on Earth. However, the unique characteristic of these satellites being situated on the same orbital path but phased 180 degrees apart holds a distinct advantage. This arrangement effectively reduces the revisit time for specific areas to 5 days. This allows for any region on Earth larger than 100km^2 being reimaged at least once every 5 days.

	S2A		S2B		
Band Number	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)	Spatial resolution (m)
1	442.7	21	442.3	21	60
2	492.4	66	492.1	66	10
3	559.8	36	559.0	36	10
4	664.6	31	665.0	31	10
5	704.1	15	703.8	16	20
6	740.5	15	739.1	15	20
7	782.8	20	779.7	20	20
8	832.8	106	833.0	106	10
8a	864.7	21	864.0	22	20
9	945.1	20	943.2	21	60
10	1373.5	31	1376.9	30	60
11	1613.7	91	1610.4	94	20
12	2202.4	175	2185.7	185	20

Fig. 4. Sentinel-2 Spectral Bands

Each Sentinel-2 satellite carries a high resolution multispectral imaging instrument called the MultiSpectral Instrument (MSI) [19] which is able to sample across 13 spectral bands ranging from visual to infrared at varying spatial resolutions as shown in Figure 4 [20]. Figure 4 shows spatial resolutions ranging from 10m per pixel for bands in the visible wavelength to up to 60m per pixel.

2.2 Vegetation Indices

Due to the spatial resolution of the available spectral bands in Sentinel-2 identifying crop types based purely on spatial features in the visible range is not possible. With a resolution of 10m per

pixel, large areas of crops will appear as the average colour over the $10m^2$ area. As a result, a number of techniques will have to be applied to extract more information such as with the use of vegetation indices. A vegetation index is a value obtained from a mathematical formula derived from remote sensing data, typically obtained from satellites or aerial imagery, to assess the health, density, and vitality of vegetation in a specific area.

The vegetation index uses the reflectance of two or more spectral bands collected from remote sensing to enhance various attributes of the vegetation, such as its structural, biochemical and plant physiological / plant stress. Biomass of green leaf, fractional cover, leaf area index (LAI) and photosynthetically absorbed active radiations are some of the examples which can be included in structural properties. Biochemical properties includes pigments (Chlorophyll, Anthocyanin, and Carotenoids), water content, nitrogen rich components, structural material like lignin and cellulose etc. Change in chlorophyll content, state of xanthophyll, moisture content are refers to plant stress [21].

Many researchers have developed vegetation indices, for determining vegetation cover and biochemical properties by using spectral data, which commonly are a linear combination of reflectance received in RED region and Near Infrared region (NIR) [21]. This paper utilises a number of vegetation indices such as the Normalised Difference Vegetation Index (NDVI) and the Enhanced Normalised Vegetation Index (ENDVI). NDVI is a widely used vegetation index can can be seen in Equation 1.

$$NDVI = (NIR - Red)/(NIR + Red) \tag{1}$$

Where NIR is near infrared and Red is red reflectance. The wavelength for NIR is between 700 and 1400μm and the wavelength of Red is between 625 and 700μm. The equation calculates the ratio of the two wavelengths for every pixel and returns a value between -1 and +1 with the higher values representing areas with live, green vegetation or areas with barren rock, sand or snow for NDVI values below 0.1[5]. An example of an NDVI image can be seen in Figure 8. In addition to extracted spatial information by utilising NDVI, information can be extracted in the temporal domain. A NDVI curve can be extracted by plotting the NDVI of a crop type over time to produce a unique spectral fingerprint, as the plant starts from early growth to harvest, that can be used to identify unlabelled crops by comparing the characteristics of the curves with other similar NDVI curves. The figure below demonstrates the variations in the NDVI curves and how features of the curve can be used to discern the crop type [22] in Figure 5.

In Figure 5, we can observe the Normalised Difference Vegetation Index (NDVI) for meadows and wheat, depicted by the orange and blue lines, respectively, plotted over the course of a year. This

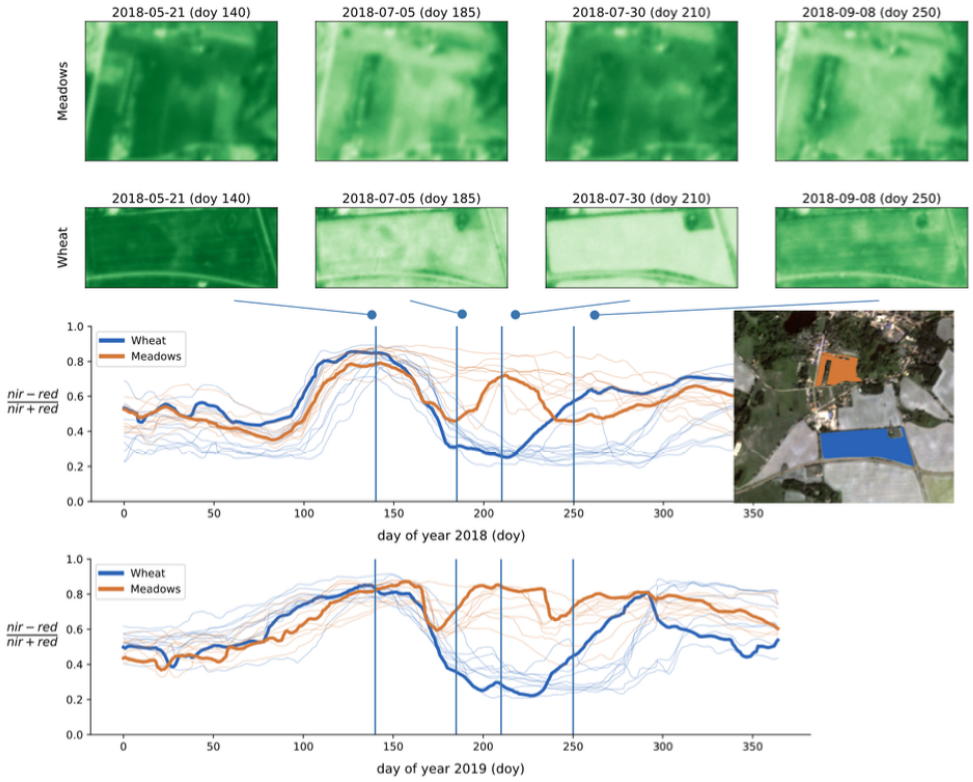


Fig. 5. NDVI Graph

graph reveals distinctive patterns in the spectral fingerprint of these two crop types, allowing for their differentiation based on their NDVI characteristics. NDVI is just one of a number of many commonly used vegetation indices which each contribute to finding characteristics of remote sensing data that might be unique to specific a vegetation index. Another example of this would be the Enhanced Normalised Difference Vegetation Index (ENDVI) which is an iteration of NDVI designed to address some of its original limitations, such as its sensitivity to atmospheric conditions and background soil reflectance, which has found application in agricultural monitoring[23, 24].

$$ENDVI = G * (NIR - Red)/(NIR + C1 * Red - C2 * Blue + L) \quad (2)$$

$$ENDVI = 2.5 * (NIR - Red)/(NIR + 6 * Red - 7.5 * Blue + 1) \quad (3)$$

Equation 2 displays the ENDVI equation where G, C1, C2 and L are constants which depend of the remote sensor. The formula shown in Equation 3 contains the constants required for ENDVI output from the Sentinel-2 remote sensor.

2.3 Deep Learning

Deep learning is a subsection of machine learning that involves training artificial neural networks to perform tasks by simulating aspects similar to human brain. Deep learning models consists of multiple layers of nodes that process and transforms data. Due to the capabilities of deep learning, it has gained a large interest in recent years and is widely applied in various application areas like healthcare, visual recognition, text analytics, cybersecurity and more[25]. A deep learning model typically follows the same processing steps as a conventional machine learning model, which consists of preprocessing steps, such as preprocessing, deep learning model building and then training, validation and testing. However, the primary difference of deep learning from other machine learning modelling is in the feature extraction stage as it is largely automated in deep learning.

When using a deep learning model there are some key features and dependencies that must be considered, for example, deep learning's dependency on data. Compared to standard machine learning algorithms, the deep learning model is typically dependant on a large amount of data to build a data-driven model for a particular problem domain. Deep learning models tend to perform poorly when the data volume is small [26], as a result, deep learning models tend to have large computational costs while training models models with large datasets and is dependant on high-performance machines with GPUs than standard machine learning methods [27].

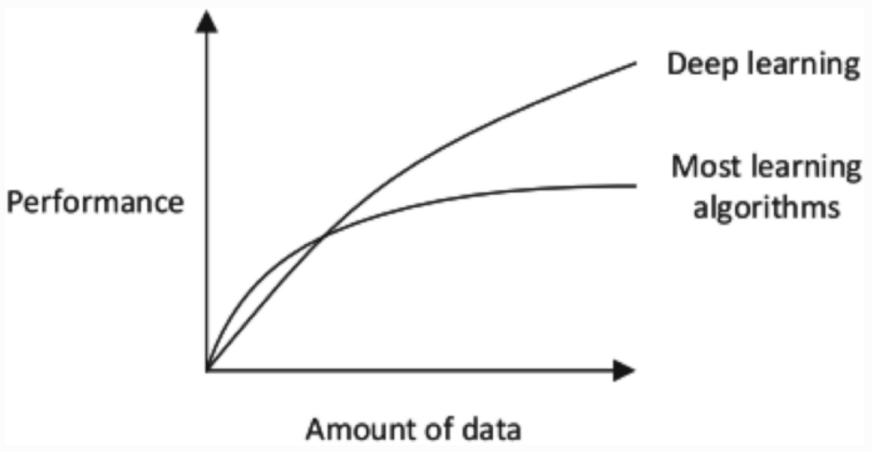


Fig. 6. Deep Learning performance Comparison

One of the primary differences between deep learning and regular machine learning is how the performance of the models increase with a larger dataset as shown in Figure 6 which demonstrates deep learning's ability to increase in performance when the amount of data grows exponentially.

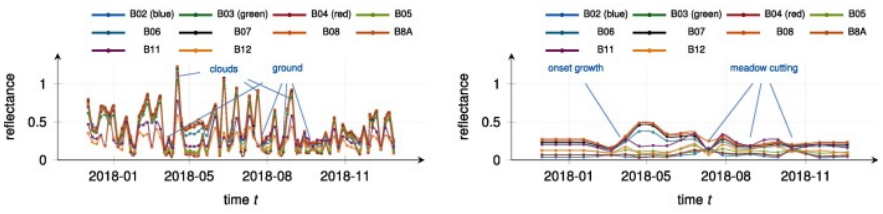
2.3.1 Dimensionality

When constructing a deep learning model, several critical factors, including dimensionality, demand careful consideration. Dimensionality, in this context, refers to the quantity of features or attributes present in the input data, and effectively managing dimensionality holds utmost importance. It directly influences the model's complexity, computational requirements, and the quality of outcomes. A key concern that arises is the Curse of Dimensionality, which presents that as the number of dimensions in a dataset increases, the task of analysing the data becomes progressively more challenging. This challenge arises because, in high-dimensional spaces, data points tend to be sparsely distributed, rendering the identification of meaningful patterns or relationships between features an arduous task.

In the specific context of this paper's dataset, the data encompasses 13 spectral bands or dimensions for each image, resulting in 72 data points throughout the year. This high dimensionality is likely to lead to suboptimal performance. To address this issue, a range of techniques can be employed to reduce the dataset's dimensions. For instance, feature selection involves carefully choosing a subset of data that retains the most pertinent information while discarding less distinctive attributes. Notably, selecting a more compact dataset not only mitigates dimensionality-related challenges but also lowers computational overhead.

3 RELATED WORK

Due to the importance of crop classification, for yield prediction and food security estimations, many techniques have been developed in the past to achieve this goal. However, due to computational limitations earlier attempts have been restricted to how the data is utilised such as using feature extraction from time series [28] such as in Figure 7.



(a) raw time series of a meadow parcel

(b) preprocessed time series of a meadow parcel

Fig. 7. Time Series Feature Extraction

Reducing the dataset into a time series can indeed be advantageous in terms of dimensionality reduction and computational efficiency. However, this transformation comes with trade-offs, primarily the loss of spatial features and, in some cases, certain temporal features. This reduction in data

can potentially lead to issues like underfitting, as the model may lack the necessary information to capture meaningful patterns adequately.

To address these challenges and enhance the model's performance in crop classification, several vegetation indices are commonly incorporated into the dataset, as demonstrated in the referenced study [28]. These vegetation indices, including NDVI (Normalised Difference Vegetation Index), NDWI (Normalised Difference Water Index), BI (Brightness Index), IRECI (Integrated Ratio-based Evaluation of Crop Indices), and EVI (Enhanced Vegetation Index), are integrated into the time series data. These indices offer valuable information about vegetation health, density, and stress levels over time, compensating for the loss of spatial details.

By including these vegetation indices, the time series data is enriched with additional features, providing the model with more comprehensive information for crop classification. This augmentation helps mitigate the risk of underfitting and contributes to more accurate and robust classification results.

With the recent improvements in deep learning [29], improvements to the techniques used for crop classification through the usage of deep learning [22]. In this techniques such as Convolutional Neural Networks [30] and Recurrent neural networks such as LSTMs [31]. In some studies [32], both CNN and RNNs are used to leverage both the spatial and temporal domain which was found to produce the most accurate results. A common limiting factor across these studies were mostly with the datasets. To produce a data, publicly available data is required from satellites such as Sentinel-2 which is limited by the spatial resolution which also needs to be paired with labels. The first limitation comes from the limited numbers of labelled crop areas as the information needs to be obtained from government surveys and be manually assigned to field areas. Furthermore, due to the limitations of publicly available remote sensing data, many other factors affect the data, ranging from limited resolution to cloud obfuscation. Some studies have found an alternative method to address this by combining data from multiple satellites, such as Sentinel-1 and Sentinel-2 in the Plant Fusion dataset, to obtain declouded data [22].

4 METHODOLOGY

4.1 Dimensionality Reduction

The dataset produced by remote sensing equipment output multi-spectral image stacks comprised of information from the spectral and electromagnetic range. Therefore, the images from remote sensing can contain more information in typical images taken in visible light spectrum beyond just

colour and texture information. The dataset used in this project is taken from the Sentinel-2 satellite program which produces data across 13 spectral bands to classify a number of crop types, although the dataset used contains 12 spectral bands. Each image stack or channel can be used for training a convolutional neural network but the dimensions of all the input features could result in increased computational complexity. Furthermore, with a high number of dimensions, the amount of data required to sample the image area would also increase exponentially with the increase in dimensions. This could result in sparsity, making it difficult for a deep learning model to find patterns and relationships in the dataset during the training. To lessen this issue, the number of dimensions in the spectral bands can be reduced though techniques such as feature selection.

4.1.1 Feature Selection

Feature selection is a method in machine learning that involves selecting a subset of the most relevant and informative features from the original set of features in the dataset. Due to the limited sample size of the dataset, features selection of the spectral bands is required to reduce the number of input dimensions. In previous research, Sentinel-2 has been employed for crop classification, with some studies making use of most of the accessible spectral bands, including up to 9 of them[33]. Due to the limited number of samples in the dataset, a reduced number of features were picked in order to mitigate issues caused by over dimensionality. 5 bandwidths (NIR, Red, Green, Blue, SWIR) were selected from the original feature space.

In addition to the subset of spectral bands selected from the original feature space, a number of vegetation indices were used to create additional image stacks to be appended to the selected subset of the bandwidths. From preliminary studies, the most commonly used vegetation indices are Normalised Difference Vegetation Index (NDVI) and Enhanced Normalised Vegetation Index (ENDVI) for crop classification. In the early stages of testing, a number of vegetation indices were trialled but due to overfitting, the number of additional layers were limited therefore NDVI was selected to be the sole supplementary layer. Figure 8 shows depicts the NDVI layer for a farm area containing wheat next to an image created using the visible bands.

4.2 Image Augmentation

In order to increase the size and the diversity of the training set, image augmentation can be performed. Image augmentation is fundamental in computer vision and deep learning and plays a crucial role in enhancing the performance of machine learning models particularly in medical [34] and agricultural applications [35]. An example of this is the application of transformations to the original images to create new training samples. Image augmentation holds significant importance

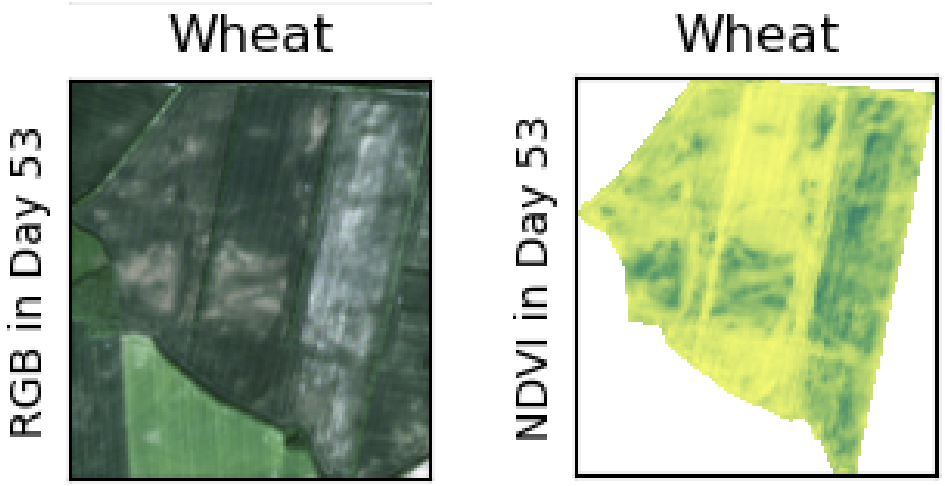


Fig. 8. Visible color and NDVI comparison

for multiple reasons. Firstly, it addresses the challenge of dataset size. In many machine learning applications, having a sizeable and diverse dataset is critical for model training and are crucial factors that significantly impact the performance of models. larger and more diverse dataset provides models with a broader range of examples, enabling them to learn more robust and generalised features. However, collecting such data, rich in real-world variations, can be a challenging and resource-intensive task [36]. Image augmentation circumvents this issue by artificially expanding the dataset size through diverse transformations.

Furthermore, image augmentation contributes to improved model accuracy to generalise it across other similar datasets. By exposing a model to a variety of augmented images during training, it becomes more capable of generalising its learned features to unseen, real-world data, resulting in better model performance on test data. Additionally, image augmentation reinforces a model's robustness in dealing with variations commonly encountered in real-world scenarios. Real-world images can vary due to factors like changes in lighting conditions, varying angles, and occlusions such as cloud coverage. Augmenting the training data with variations that closely resemble these real-world scenarios equips models to be more resilient and capable of handling such variations during its application.

Some commonly employed image augmentation techniques include flipping, rotation, scaling, noise injection, sharpening, translating, cropping and adjusting the contrast[37]. These transformations introduce diversity into the training dataset, enabling models to learn invariant features and adapt to the plethora of conditions they may encounter in the real world.

Figure 9 displays the crop type distribution of the dataset used in this study which exhibits a notable class imbalance. Class imbalance refers to a situation where the distribution of classes within

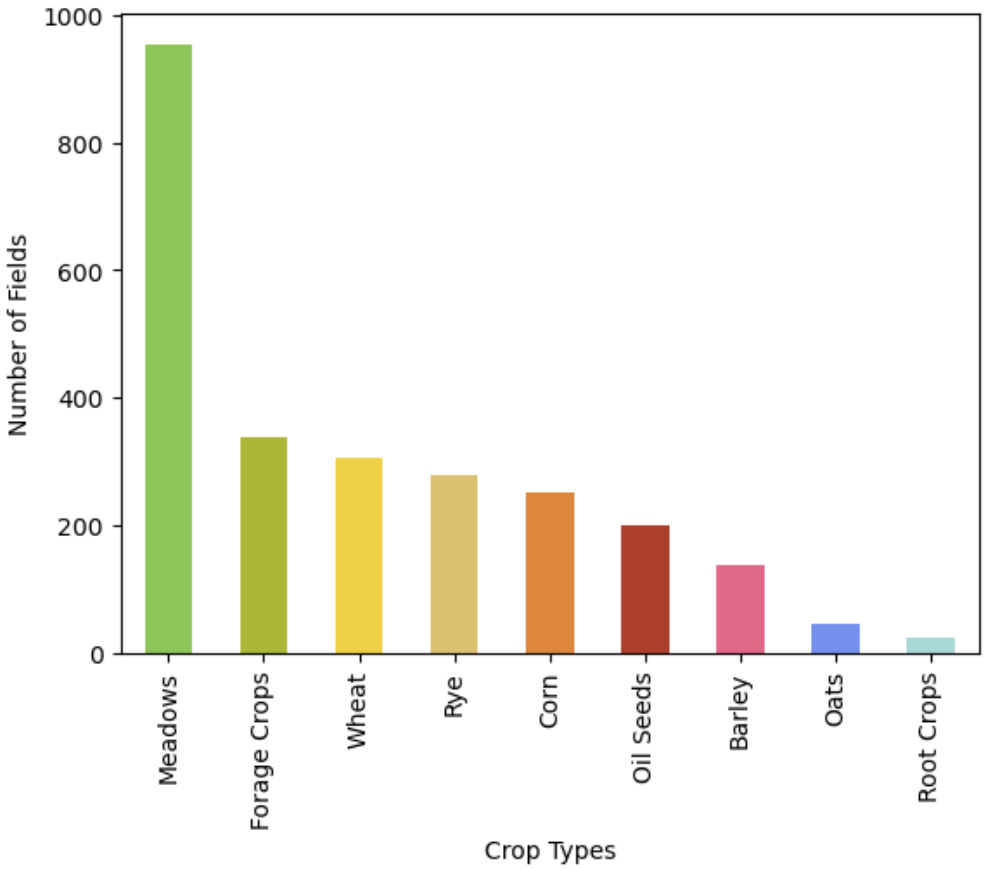


Fig. 9. Distribution of Crop Types

a dataset is highly skewed, resulting in one or more classes being significantly underrepresented compared to others. This is particularly evident in the difference in frequencies between the minority classes (Root Crops and Oats) and the majority classes. This class imbalance presents an opportunity to implement data augmentation to reduce the class imbalance. Initially, the meadow instances were reduced in number, and the labels of minority classes were duplicated until there were a total of 400 instances for each crop type. Additionally, the image stacks underwent random transformations, including horizontal and vertical flips, rotations, cropping, and the addition of Gaussian noise.

4.3 Deep Learning Model Design

Deep learning has emerged as a powerful tool across various fields, enabling the extraction of intricate patterns and insights from complex data. Effective model design is at the core of it's success and effectiveness, influencing the model's capacity to capture meaningful information and to make accurate predictions. Examples of deep learning architectures include Convolutional

Neural Networks (CNN) and Long Short-term Memory (LSTM). A CNN is commonly used for tasks related to computer vision and used to solve difficult image-driven pattern recognition tasks such as image classification[38]. CNNs are designed to automatically and adaptively learn spatial hierarchies of features from input data, making them effective at capturing visual patterns and structures and consist of convolutional layers, pooling layers, and fully connected layers. A LSTM is a type of recurrent neural network (RNN) architecture, designed for sequence prediction and sequence labeling tasks [39]. By combining the convolutional neural network and long short-term memory to create CNN LSTM, a model can be designed for sequence predictions in images and videos. The CNN LSTM model utilises the CNN layers for feature extraction in the spatial domain the the LSTM layers to extract features in the temporal domain. Figure 10 displays the CNN LSTM architecture designed for this paper.

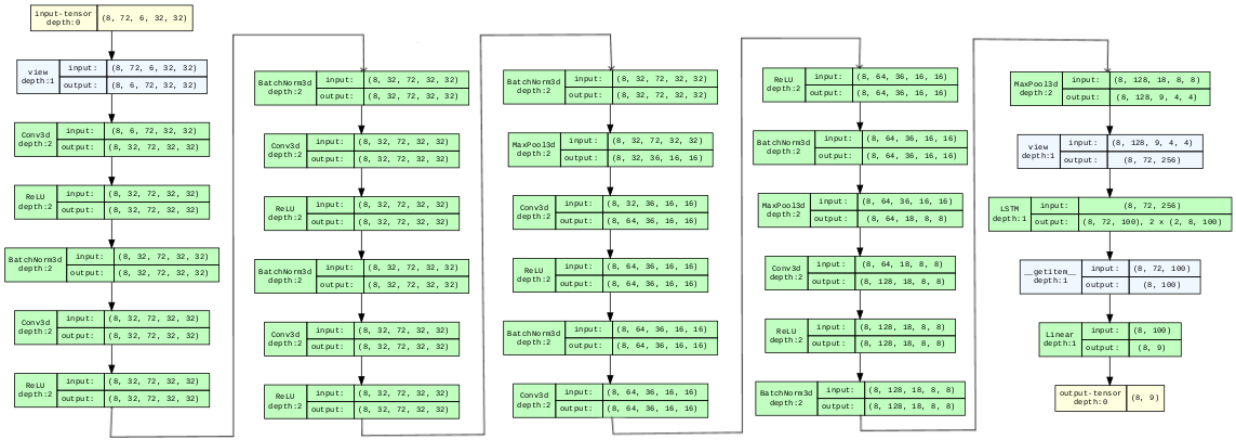


Fig. 10. CNN LSTM Architecture

The CNN layers that handle feature extraction in the spatial domain, consists of: convolutional layers, pooling layers, batch normalisation and ReLU layers.

4.3.1 Convolutional Layer

In the case of figure 10, 3D convolutional layers are utilised. These layers apply a convolutional operation to the input, then passing the output to the following layer. Each convolution converts all in pixels in the receptive field into a single value and with each repeated application of the same filter results in a map of activation's known as a feature map, indicating the location and strength of a feature from the input. 3D convolution was applied to the input, the mathematical operation can be seen in equation 4.

$$\text{out}\left(N_i, C_{\text{out}_j}\right) = \text{bias}\left(C_{\text{out}_j}\right) + \sum_{k=0}^{C_{\text{in}}-1} \text{weight}\left(C_{\text{out}_j}, k\right) \star \text{input}\left(N_i, k\right) \tag{4}$$

In equation 4,

N - Batch Size.

C - Number of images in the image stack / length of data sequence.

D - Dimensions of input. In this case, the number of selected spectral bands used in the input.

H, W - Height and width of the input images, measured by the number of pixels.

3D convolutional is used to capture all the information in the spatial domain. Each image used in the model consists of 32x32 resolution images of which there are multiple of due to information being captured in multiple spectral bands.

4.3.2 Pooling Layer

Pooling layers are used in CNNs to reduce the spatial dimensions of the input feature maps while preserving the depth. The input features map is turned into a set of pooling regions by dividing the input into non-overlapping regions. The pooling layer then operates on each feature map to create a new set of pooled features maps. In figure 10, max pooling layers are used in the CNN LSTM model. Max pooling works on the pooling regions and retains the maximum value within each region while discarding the rest. This reduces the spatial resolution of the feature maps while preserving the most important information. The mathematical operation for 3D max pooling is shown below in equation 5.

$$\begin{aligned}\text{out}\left(N_i, C_j, d, h, w\right) = &\max_{k=0,\ldots,kD-1} \max_{m=0,\ldots,kH-1} \max_{n=0,\ldots,kW-1} \\ &\text{input}\left(N_i, C_j,\ \text{stride}[0] \times d + k,\ \text{stride}[1]\ [1]h + m,\ \text{stride}\ [2] \times w + n\right)\end{aligned} \tag{5}$$

In equation 5,

N - Batch Size.

C - Number of images in the image stack / length of data sequence.

D - Dimensions of input. In this case, the number of selected spectral bands used in the input.

H, W - Height and width of the input images, measured by the number of pixels.

4.3.3 Batch Normalisation

Batch normalisation layers are a technique used in deep neural networks to normalise the inputs to a layer and are used to stabilise and accelerate the training of models. They work by normalising the activation's of a layer by adjusting their mean and standard deviation within mini-batches during

training. Batch normalisation reduces the internal covariate shift, helping the network converge faster and enabling the use of higher learning rates. It also acts as a regulariser to reduce the risk of overfitting.

$$y = \frac{x - \mathrm{E}[x]}{\sqrt{\mathrm{Var}[x] + \epsilon}} * \gamma + \beta \tag{6}$$

4.3.4 ReLU Layer

A Rectified Linear Unit (ReLU) is a non-linear activation function that is used in multi layer neural networks and is conventionally used as an activation function for the hidden layers in a deep neural network [40]. The ReLU activation functions is applied to the output of each neuron in the network and will introduce non linearity to the model by outputting the input if the value is positive or output zero if the value is negative resulting in negative values in the filtered image being removed.

4.3.5 Long Short-Term Memory Layer

Long Short-Term Memory layers are a type of recurrent neural network (RNN) layer that is able to learn long-term dependencies between time steps in a time series and sequence data. LSTMs achieve this by incorporating memory cells and gating mechanisms. Each LSTM unit contains memory units which have the ability to store information over a number of time steps which allows the LSTM layer to capture long-term dependencies in the time series data. The gating mechanisms incorporated in the LSTM enables the layer to control the flow of information, prevent vanishing gradients, and make accurate predictions based on sequential input.

4.3.6 Fully Connected Layer

Fully Connected (FC) layers in a neural network are layers where all the inputs from one layer are connected to all the activation units of the following layer which results in a dense and fully interconnected architecture. Each neuron in a FC layer applies a linear transformation to the input vector using a matrix of weights, which as a result, introduces non-linearity to the neural network allowing the network to model complex, non-linear relationships in the data.

4.4 Classification Strategies

In image classification, various techniques can be employed during the classification stage. Two prominent approaches are multiclass and binary classification. In multiclass classification, the primary goal is to assign data samples into multiple classes, ensuring that each sample belongs to

one and only one category. On the other hand, binary classification, specifically the One-Versus-All method, focuses on distinguishing a chosen class as the positive class while considering all other samples as the negative class.

4.5 Dataset Format

The dataset used is obtained from the Food Security Challenge by AI4EO[41] which covers an area of land near Brandenburg in Germany with high-quality cadastral data on field boundaries and crop types as ground truth input. The purpose of the dataset over Germany is to test the reusability of models for crop identification from one growing season to the next therefore, the dataset spans over 2 years from 2018 to 2019. Due to the revisit time of Sentinel-2, over the 2 years, 144 length time series of images was collected over the Brandenburg area which contains over 2500 field areas which are each labelled for the 2018 year. 9 crop types were labelled and can be seen in Figure 11, however, the labels are only valid for 2018 so that dataset was cut in half to contain just 2018 data or the first 72 datapoints.

Crop ID	Crop Type
1	Wheat
2	Rye
3	Barley
4	Oat
5	Corn
6	Oil
7	Root Crops
8	Meadows
9	Forage Crops

Fig. 11. Crop ID table

When developing machine learning models, it is essential to divide the dataset into training, validation and testing datasets to train, fine-tune, and evaluate a machine learning model in a systematic and unbiased manner, ensuring that it can make accurate predictions on new and unseen data. The training data is comprised of the most significant portion of the original dataset and is used to to teach the model to recognize patterns and relationships within the input data to associate which labels correspond with the various crop types in the dataset. The validation dataset is employed during the model training process to fine-tune hyperparameters, assess model performance, and prevent overfitting. By evaluating the model on the validation data, necessary adjustments can be

made to optimise its accuracy. The test dataset is reserved for the final evaluation of the trained model. It is kept separate from the training and validation datasets to ensure unbiased assessment. After the model has been trained and validated, it is subjected to the test data to gauge its performance in real-world scenarios. The results obtained from the test set provide an accurate measure of how well the model can generalize to new, unseen data.

5 RESULTS AND DISCUSSION

5.1 Image Preprocessing

The dataset used in this study comprised of data spanning two years, 2018 and 2019, and included crop type labels for each individual field area for 9 types of crop. However, the original intention for the dataset was to ensure that a model can handle a domain shift to a different year, therefore, only the labels were only valid for the 2018 remote sensing data [42]. As a result, the image stacks were trimmed down from 144 images to 72, for the dataset to only contain the images from 2018.
Furthermore, many of the images in the data are affected by the cloud coverage or the shadows, which both account for half the the images in the stack. An attempt to augment the data was made to recover non-cloud affected data from the image stacks. Figure 12 displays clouded data of the same field area displayed in figure 8 but for the following Sentinel-2 revisit.

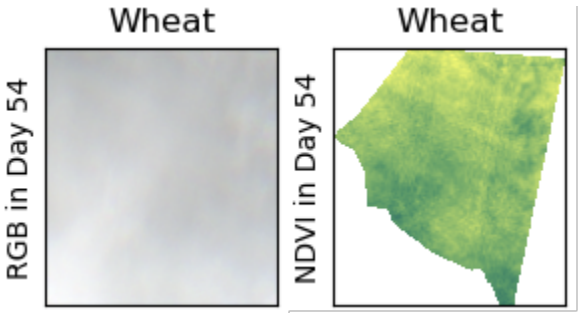


Fig. 12. Cloud Obscured Field Area

Sentinel-2 remote sensing data includes a cloud mask that measures the levels of cloud coverage at a 60m resolution. An example of the cloud mask can be seen in Figure 13 which displays the clouds on a gradient from white to black with the latter representing the cloud on a scale of 0-255.

The effects of the cloud coverage over a certain field can be seen in figure 14.

Figure 14 displays the mean cloud coverage over a randomly selected field area. The values above 200 suggests almost total cloud coverage and the lower values suggests that the field area is partially covered by clouds. Due to the difference in cloud coverage, different techniques will need to be

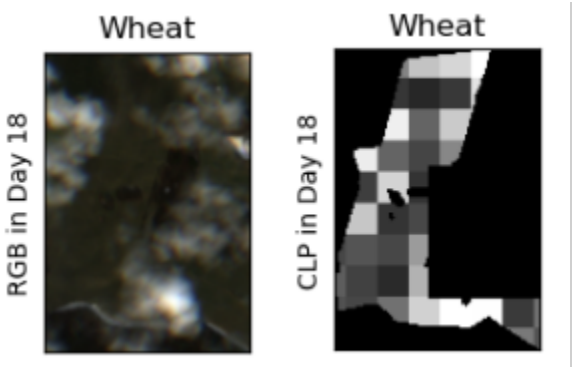


Fig. 13. Sentinel-2 Cloud Mask

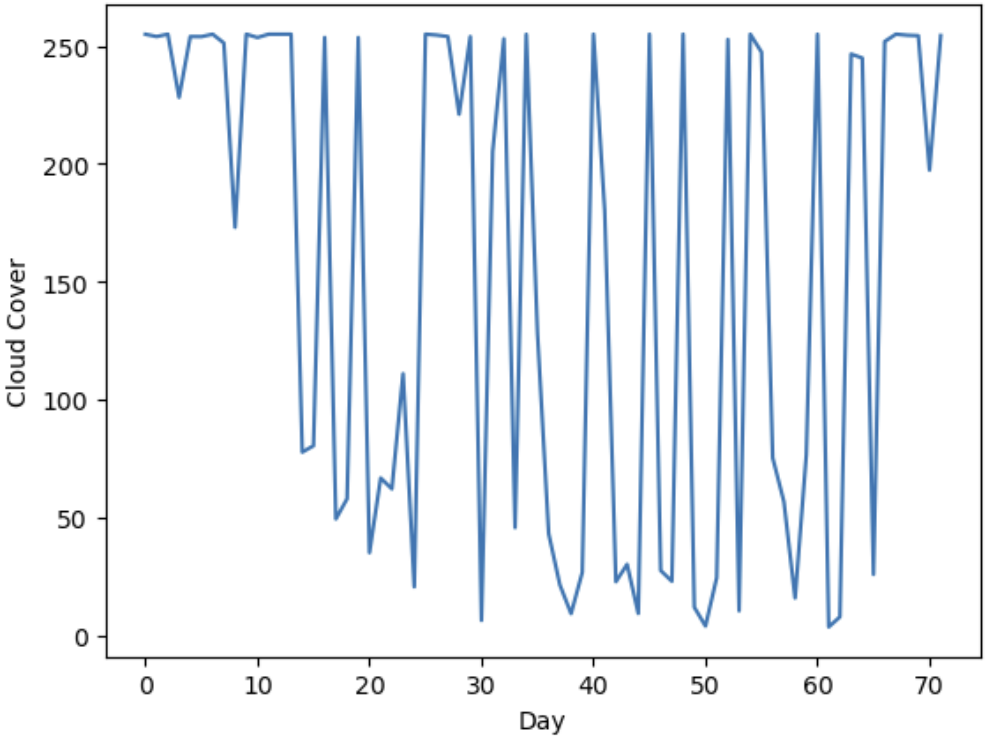


Fig. 14. Mean Cloud Coverage

used. For the days with partial coverage spatial interpolation can be implemented and for days with complete cloud obfuscation temporal interpolation can be utilised. Due to the nature of the data, in an ideal situation, the spectral data across a field area would be uniform which would also result in each datapoint being a single value rather than a 32 x 32 image. Therefore, by reducing the data to just the median pixel value for each pixel a 2D representation of the dataset can be created with the second dimension being the multiple spectral bands. During this process, spatial interpolation can be implemented by selected the mean pixel values from areas with minimal or low cloud coverage

which can be found by analysing the cloud mask. Following the transition to the 2D data structure, temporal interpolation can be introduced by utilising data from before and after dates with low cloud coverage to approximate the pixel values. This solution was trialled for the dataset however due to the high cloud coverage, as show in Figure 14, limited options were given to temporally interpolate the data which resulted in negligible improvements from the model.

5.2 Image Augmentation

In the data, one of the most significant challenges encountered is class imbalance, as depicted in Figure 9. This disparity is particularly pronounced in certain classes, where the numbers differ by an order of magnitude, such as root crops and forage crops. To address this critical issue, multiple approaches were explored.

5.2.1 Class Weights

Initially, class weights were considered. These weights are determined inversely proportional to the frequency of each class. This means that underrepresented minority classes receive a greater weight, while overrepresented majority classes receive a lower weight. However, in the final model, a resampling technique was chosen.

5.2.2 Resampling

This chosen technique involves both oversampling of minority classes and undersampling of majority classes. To implement this method and alleviate class imbalance, it's crucial to select a target number of samples. Striking the right balance in the number of samples is essential. Excessive computational cost should be avoided, while ensuring that there are enough samples to retain valuable information, diversity, and training data for the model.

In this context, it was decided to set the target number of samples at 400 for each crop type. This choice aims to strike a balance between computational efficiency and the need to preserve information and diversity within the samples used for training. By resampling in this manner, the model can better learn from the data, account for class imbalances, and make more accurate predictions across all crop types.

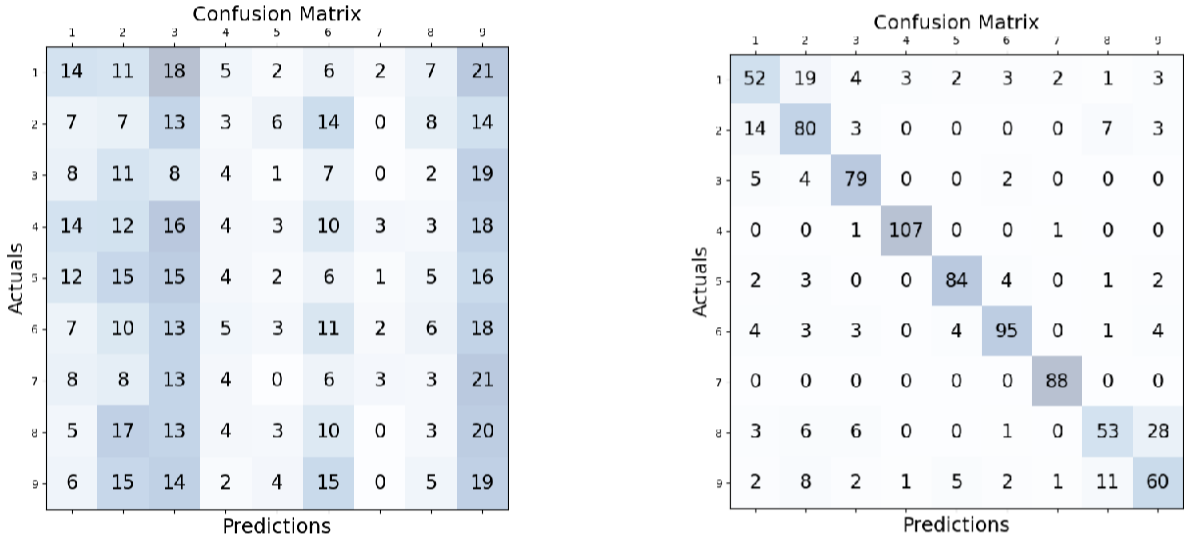
In addition to mitigating class imbalance, a variety of image transformations were incorporated into the dataset. These transformations encompassed image rotation, flipping (both horizontally and vertically), random cropping, and the introduction of random noise. These augmentations were employed to diversify the dataset and expose the model to a broader spectrum of variations. Rotation allowed the model to learn from objects at different angles, while flipping simulated

varying viewpoints and mirror reflections. Random cropping encouraged the model to focus on different regions within each image, enhancing feature robustness. Crops were reduced to 32x32 pixel resolution for the mode. Furthermore, the injection of random noise prepared the model to handle noisy or less-than-ideal input conditions, contributing to its overall adaptability and performance in real-world scenarios.

Additionally, the dataset required preparation for both binary and multiclass classification tasks. For the multiclass model, a balanced dataset was created with 400 samples for each crop type. However, in the binary classification model, a one-versus-all (1 v all) approach was adopted. In this scheme, for each specific crop type under consideration, the dataset consisted of 400 samples from that primary crop type, along with 400 crop samples randomly selected from the remaining eight crop types. This approach facilitates the training of binary classifiers for each individual crop type against the rest, ensuring that the model can effectively distinguish between the target crop and all other crops in the classification task.

5.3 Evaluation

This section assesses several models in comparison to the final model, as illustrated in Figure 10. All models underwent training using the same dataset, employing an early stopping mechanism set to terminate training after 15 epochs.



(a) Underfitting Confusion Matrix

(b) Final Confusion Matrix

Fig. 15. Confusion Matrix Comparison

Figure 15 provides a visual comparison between two models: one after the application of data augmentation to mitigate class imbalance (on the left) and the final baseline model (on the right). The correspondence between each crop ID and its corresponding crop type can be found in Figure 11.

To address the class imbalance, a combination of oversampling and undersampling techniques was applied, ensuring an equal number of samples for each class within each crop type. However, when examining the relationship between the class distribution chart in Figure 9 and the confusion matrix results in Figure 15, it becomes evident that there is a discrepancy. In some cases, underrepresented classes are predicted more frequently than the majority classes. This discrepancy suggests that the model might be overfitting the data.

One potential explanation for this phenomenon, especially after oversampling to address class imbalance, is that there were initially too few samples of the minority classes. Consequently, even after applying transformations, the resulting samples may remain very similar, leading to an overemphasis on these classes during training. Further investigation into model complexity and potential overfitting is warranted to understand and address this issue effectively.

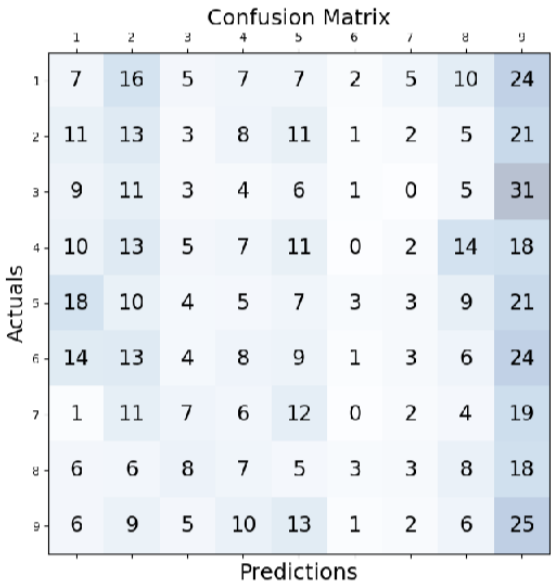
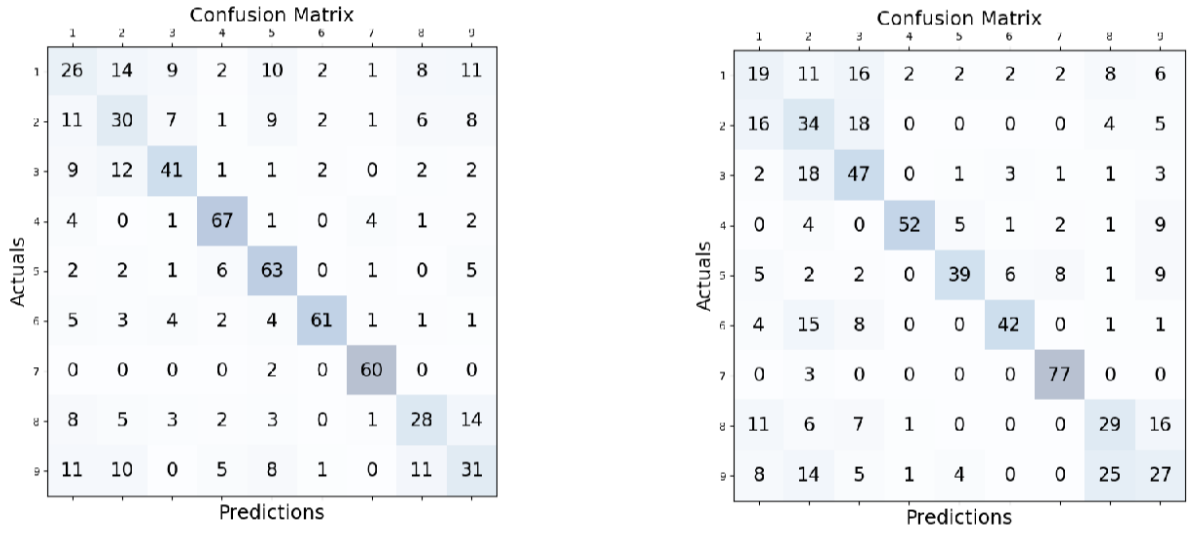


Fig. 16. Improved Model Confusion Matrix

Figure 16 presents the confusion matrix for a model with an increased number of convolutional layers, aimed at reducing the underfitting observed in Figure 15. The adjustments made to the model have led to a more balanced distribution of predictions across each class, but certain trends persist, particularly noticeable in column 9, which pertains to "Forage Crops."

One potential explanation for this ongoing trend lies in the labeling of the dataset, where similar crop types have been grouped into a single category. This grouping introduces additional complexity for the model, as it must learn to distinguish between various subtypes within a broader category. For instance, in column 9 of Figure 16, the crop type is designated as "Forage Crops," which encompasses a range of species and crops primarily cultivated for livestock feed. This category includes various grasses, oats, barley, rye for hay production, and even maize (corn) for silage. The diverse range of crops within this category can result in some data overlap, as different species share similar characteristics. Consequently, the model may encounter challenges in accurately distinguishing between these closely related subtypes, contributing to the observed trends in the confusion matrix.



(a) 3 Convolutional Layers

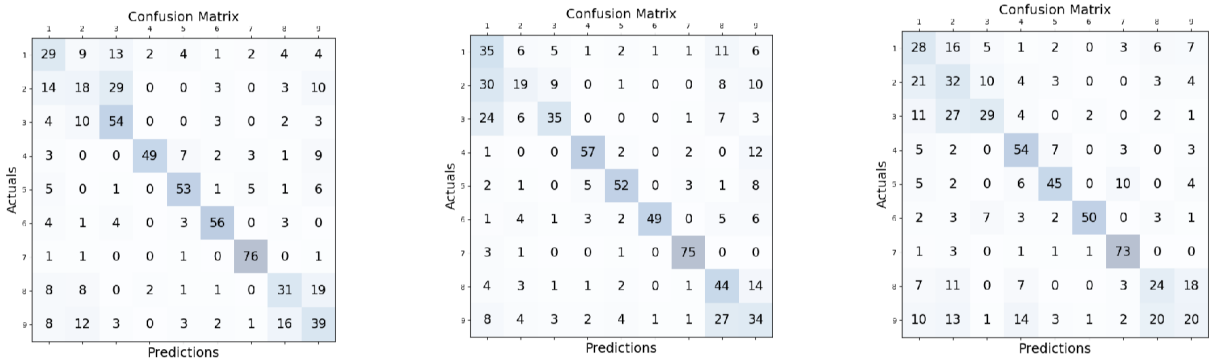
(b) 6 Convolutional Layers + ENDVI

Fig. 17. Convolutional Layer Comparison

To enhance the model's predictive capabilities and address the challenges in correctly classifying crop types, an increase in complexity was deemed necessary. This increase in complexity was achieved by augmenting the number of convolutional layers. Figure 17 illustrates the outcomes of this complexity enhancement, where two models are presented: one employing 3 convolutional layers and another utilising 6 layers. Additionally, the rightmost model in the figure incorporates an ENDVI (Enhanced Normalised Difference Vegetation Index) layer into the image stack.

The model with 3 convolutional layers immediately demonstrates notable improvements compared to the results depicted in Figure 16. A distinct diagonal line in the confusion matrix becomes evident, indicating that the improved model is beginning to make correct predictions for crop types. However, there are still some inaccuracies observed in crops 1, 2, 3, and 9.

In pursuit of further improvement, the number of convolutional layers was further increased to 6, and an additional ENDVI layer was introduced into the image stack. However, subsequent testing revealed that the benefits gained from the additional channel introduced by the ENDVI layer were marginal over the 5 spectral bands and the NDVI channel. Consequently, to mitigate the risk of overfitting and reduce dimensionality, the ENDVI layer was removed from the model while maintaining the enhanced complexity achieved through the 6 convolutional layers.



(a) 1 LSTM (b) 2 LSTM (c) 3 LSTM

Fig. 18. LSTM Layer Comparison

In addition to modifying the number of layers in the Convolutional Neural Network (CNN) to enhance performance in the spatial domain, adjustments were also made to the number of Long Short-Term Memory (LSTM) layers to improve performance in the temporal domain. The outcomes of these adjustments and testing are presented in Figure 18. The train and validation losses for the 1 and 3 layer LSTM model can be seen in Figure 19 and Figure 20 respectively.

(a) Train Losses (b) Valid Losses

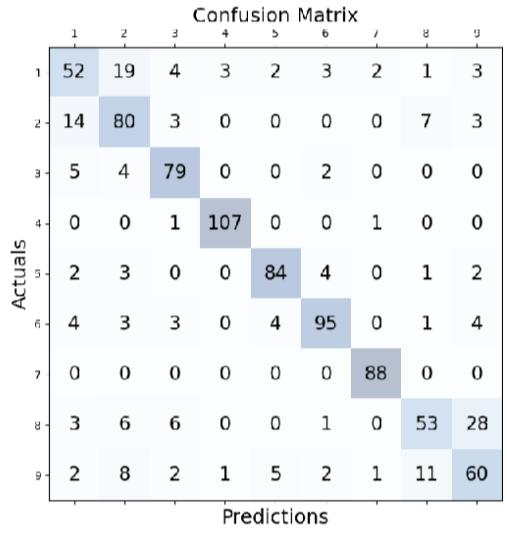
Fig. 19. 1 LSTM Losses

Upon examining the losses for Figure 19 and Figure 20, a decision to utilise 2 LSTM layers was reached. The losses in Figure 19 indicated that the validation losses consistently exceeded the training losses, suggesting that a single LSTM layer was insufficient for capturing temporal details and that the model was underfitting. Conversely, Figure 20 displayed signs of overfitting, with training losses significantly lower than validation losses, and the results degraded, as evident in

(a) Train Losses (b) Valid Losses

Fig. 20. 3 LSTM Losses

the confusion matrix in Figure 18. Moreover, employing 3 LSTM layers increased computational costs. Therefore, in pursuit of the best model, it was determined that 2 LSTM layers strike an optimal balance, effectively capturing temporal patterns without introducing excessive complexity and computational demands. In the end, the multiclass model was designed with six convolutional layers, each incorporating batch normalisation and ReLU activation functions, with a max pooling layer alternating for every other convolutional layer. These architectural choices contributed to the model's overall performance and ability to distinguish between various crop types.



(a) Baseline model - Confusion Matrix

Crop ID	Crop Type
1	Wheat
2	Rye
3	Barley
4	Oat
5	Corn
6	Oil
7	Root Crops
8	Meadows
9	Forage Crops

(b) Crop ID table

Fig. 21. Final Baseline Model

In the final multiclass model, an accuracy of approximately 0.79 was attained. The performance results are illustrated in the confusion matrix provided in Figure 21, and the corresponding mapping between crop IDs and crop types can be found on the right. Addtionally, the losses for the baseline model can be seen below in Figure 22.

(a) Training loss (b) Validation loss

Fig. 22. Final Baseline Model losses

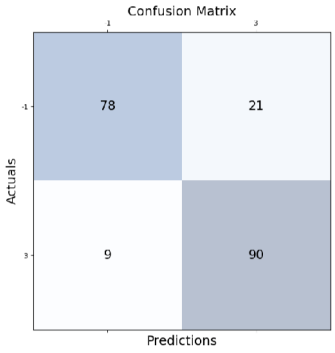
Figure 21 reveals varying degrees of success for each crop type. Notably, there are challenges in correctly classifying crop types 1 (Wheat), 8 (Meadows), and 9 (Forage Crops), with lower success rates observed. Conversely, crop types 4 (Oat), 5 (Corn/Maize), and 6 (Oil) exhibit higher levels of successful classification.

In summary, while the model's performance has shown promise with an accuracy of approximately 0.79, it is evident that further enhancements are required to make it reliably applicable for real-world scenarios. An alternative approach under consideration is the implementation of binary classification. As observed in Figure 21, the baseline model exhibits varying levels of success in predicting different crop types. To address this, a one-versus-all binary classification model can be employed to classify each crop type individually, potentially leading to improved results.

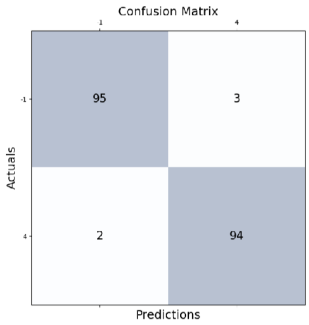
To adapt the model for binary classification, adjustments to the dataset are necessary. In the baseline model, a dataset containing 400 samples of each crop type was used throughout training, validation, and testing. However, for training the one-versus-all model, the number of samples for each crop type must be modified to ensure an equal number of samples for the target crop type and an equal number of samples from all other crops, randomly selected from those not belonging to the target crop type. This approach aims to balance the dataset and enable more focused and effective binary classification for each crop type of interest. This is achieved by altering the label data to delete crops that are not needed for the classification and by changing the crop ID of the targeted crop to 1 and the other crops to -1, a confusion matrix can be created.

Figures 23, 24, and 25 present the confusion matrices for each of the crop types, and the corresponding accuracies for each crop type are depicted in Figure 26. Upon careful examination, it becomes evident that the accuracies achieved for each crop type are comparable to the results obtained from the baseline model as shown in Figure 21.

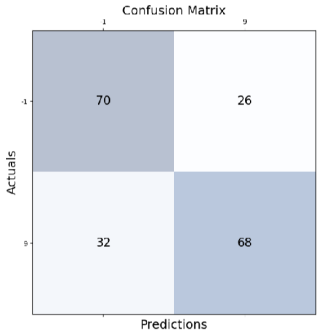
However, the key advantage of the one-versus-all (1vAll) approach is its ability to target specific crop types individually, resulting in more accurate classifications for selected crops. This finer granularity of classification enables the model to excel in distinguishing certain crop types, which might have been challenging in a multiclass classification setting. The approach empowers tailored



(a) Barley

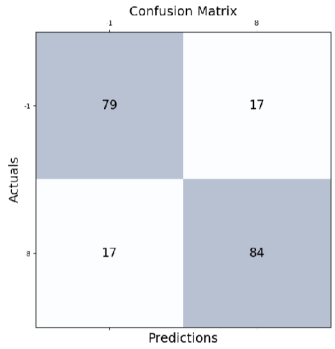


(b) Corn

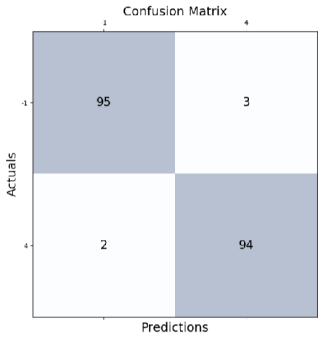


(c) Forage Crops

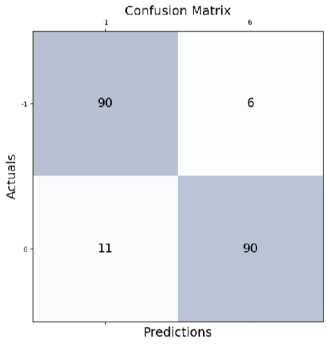
Fig. 23. One versus All comparison 1



(a) Meadow

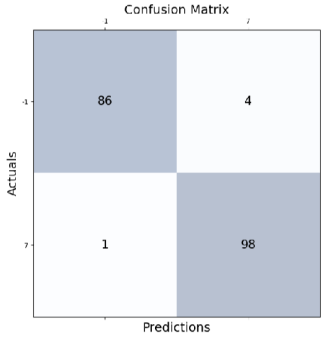


(b) Oat

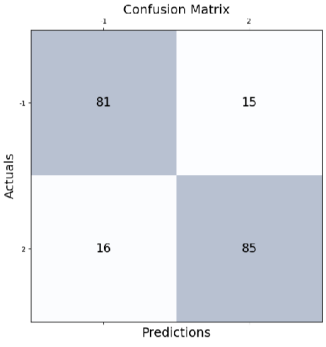


(c) Oilseed

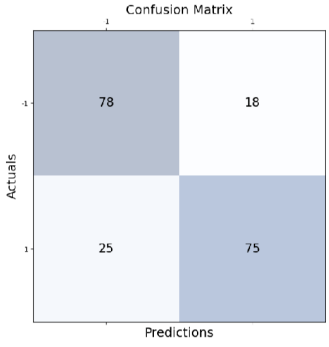
Fig. 24. One versus All comparison 2



(a) Root Crops



(b) Rye



(c) Wheat

Fig. 25. One versus All comparison 3

solutions for each crop type, enhancing overall classification performance and addressing the specific needs of different crops in the agricultural context.

Crop Type	Accuracy
Wheat	0.79
Rye	0.8
Barley	0.87
Oat	0.94
Corn	0.94
Oil	0.89
Root Crops	0.98
Meadows	0.79
Forage Crops	0.72

Fig. 26. One versus All results

5.4 Variable Time Series Length

One of the primary objectives of this project is to enable the classification of crops before the entire sowing and harvest cycle is completed. In the current baseline model, crop data from the entire year of 2018 has been utilised for training. However, for the initial trial of the one-versus-all models, it is essential to explore the impact of varying time frames. Understanding the estimated time frames for the sowing and harvesting of each crop is crucial for this purpose.

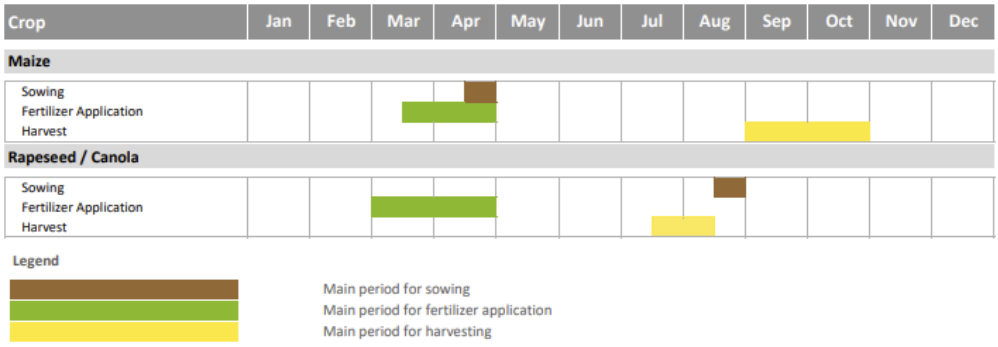
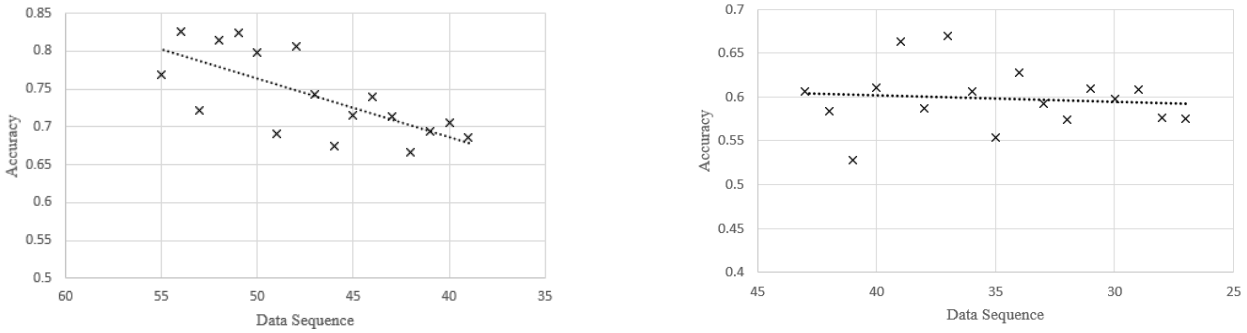


Fig. 27. Crop Harvest Timetable

Figure 27 provides valuable insights into the crop harvest timetable for Germany, as sourced from [43]. Specifically, it highlights the harvest periods for maize (corn) and rapeseed/canola (oilseed). Using this data, we can determine the relevant time frames required for crop classification. Corn and oilseed were specifically selected from the pool of nine crop classes in the baseline model. This selection was made taking into account the number of unique samples available for each crop class before data augmentation. In the initial dataset, certain crop classes, such as root crops, had a

relatively limited number of unique samples, approximately around 20. Consequently, when working with such imbalanced data, the model's accuracy might be misleadingly high, often reaching levels as high as 0.98 or more. However, this high accuracy can be attributed to the class imbalance rather than the model's true generalisation performance. To ensure a more accurate evaluation of the model's capabilities, the decision was made to focus on corn and oilseed as target classes. These crop classes were chosen because they offered a more balanced number of unique samples, allowing for a more reliable assessment of the model's performance. Analysing Figure 27, we can ascertain that corn is typically harvested around the 273rd day of the year, while oilseed harvesting occurs around the 212th day. Given the 5-day revisit time of Sentinel-2, which results in approximately 73 data points per year, we can optimise the dataset by retaining only the pertinent data points. This entails narrowing down the data to approximately 55 data points for corn and 43 data points for oilseed, focusing specifically on the relevant harvest periods for each crop. This tailored approach allows us to align the data with the critical phases of each crop's growth cycle, enhancing the precision and timeliness of crop classification.



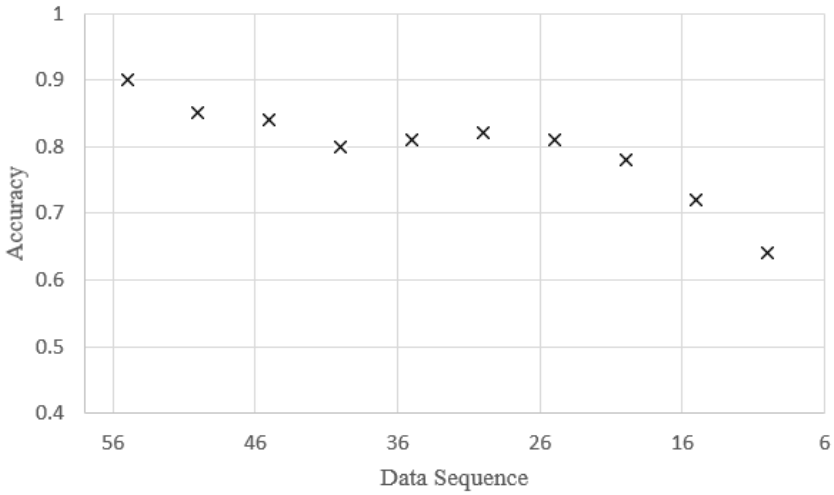
(a) Corn Time Series

(b) Oil Time Series

Fig. 28. Corn and Oil Variable Time Series Length

The data presented in Figure 28 illustrates some initial testing involving the manipulation of time series lengths. In this experimentation, the initial step was to shorten the length of the time series by removing the end section, retaining only the data within the growth period of interest. However, it's important to note that the model in use requires a fixed-length time series as input. To address this requirement, an arbitrary approach was employed, involving the repetition of the final image of the image stack until the time series reached a length of 72. While this approach provides a fixed-length input, it comes with certain limitations. Specifically, altering the ends of the time series in this manner can lead to inaccuracies in crop classification. This is because the model continues to explore features beyond the chosen end of the sequence, potentially introducing inconsistencies and misclassifications. The results presented in Figure 28 reveal a clear descending trend as the length of

the time series decreases. This trend is indicative of the sensitivity of the model to the length of the input time series.Notably, the accuracy of crop classification for corn and oil, which initially stood at 0.94 and 0.89, respectively, in the baseline tests, experiences a significant drop when the time series is trimmed down to cover only the specific growing periods of these crops. Specifically, the accuracy decreases to approximately 0.8 for corn and approximately 0.6 for oil. These findings underscore the importance of considering the temporal context in crop classification tasks. It demonstrates that the model's ability to accurately classify crops relies heavily on having access to a sufficient and relevant historical context within the time series data. Reducing the length of the time series to only include the growth periods of the crops, without using a different model, can lead to a loss of critical information and, consequently, a decrease in classification performance. This highlights the necessity of aligning the model's input data with the relevant temporal context for reliable crop classification results. Another approach was to train a new model targeted at classifying individual crops at certain periods throughout the year. Corn was chosen for this model as it represents a greater portion of the dataset at 55 revisists before harvest compared to 43 datapoints for oilseed. A new model was trained at 5 datapoint intervals, which corresponds to 25 day gaps, beginning from datapoint 55.



(a) Time Series Length

Day	Accuracy
55	0.9
50	0.85
45	0.84
40	0.8
35	0.81
30	0.82
25	0.81
20	0.78
15	0.72
10	0.64

(b) Results Table

Fig. 29. Corn Variable Time Series Length

As illustrated in Figure 29, notable improvements are evident when compared to the results shown in Figure 28. The model's performance starts with a higher accuracy of 0.9 compared to the previous 0.8. Furthermore, the accuracy in Figure 29 remains relatively consistent, with results consistently

above 0.8 until the end of the testing period. Toward the end, the results exhibit a rapid decline due to the absence of clean data caused by high cloud coverage during the early part of the year.

It is evident that training a specialized model designed for classifying individual crops during specific growth periods throughout the year yields promising results. However, it's important to note that the dataset used in this study lacks data from the earlier stages of corn growth, which could potentially lead to even higher accuracy and more stable results, particularly during the cloud-covered winter period. All the results can be found under temp_s2 file in the code.

6 CONCLUSION

This study proposes a deep learning model able to classify crop types from farm areas by exploiting the information in the spatial and temporal domain. The dataset used to train the model uses Sentinel-2 remote sensing data over a region near Brandenburg containing over 2500 field areas. This data is collected across 12 spectral bands from the beginning of 2018 to the of of 2019. The proposed deep learning model utilised information from the spatial and temporal domain using a CNN LSTM architecture to initially perform multiclass classification but was found to perform different depending on the crop type. As a result, a switch from multiclass to binary classification was made to design a model targeted at classifying specific crops to ensure the greatest accuracy. The results from these models demonstrated the limitations of the dataset, for example, the grouping of subspecies of multiple crop types into a single group such as forage crops and root crops. Another limitation of dataset class imbalance which resulted in an order of magnitude of difference between the majority and minority class which was solved by undersampling and oversampling. Finally, some tests of varying the length of the time series was experimented on where it was found to perform well by training a new model targeted at classifying a specific crop through set time periods through the year. The greatest limitation of the experiment is the lack of data from the earlier stages of growth due to the dataset beginning at the start of 2018 but the planting period begins around September 2017. With the data from the previous year, I believe the average accuracy will be greater and the decline in accuracy when only including the start of the time series will not be as present. Additionally, all the data used to create the deep learning model utilised dataset from Brandenburg so it is likely that this model will only provide good results in regions areas.

References

[1] Foley J., Ramankutty N., Brauman K., and et al. *Solutions for a cultivated planet.* Nature, 2011.

[2] SAURAB BABU. What are improper agricultural practices doing to our land?, 2016. Accessed September 28, 2023. https://eco-intelligent.com/2016/11/05/what-are-improper-agricultural-practices-doing-to-our-land/.

[3] Olaniyi A. Ajadi a, Jeremiah Barr a, Sang-Zi Liang a, Rogerio Ferreira a, Siva P. Kumpatla a, Rinkal Patel b, and Anu Swatantran. *Large-scale crop type and crop area mapping across Brazil using synthetic aperture radar and optical imagery*. Elsevier B.V., 2021.

[4] Gideon Okpoti Tetteh, Alexander Gocht, Marcel Schwieder, Stefan Erasmi, and Christopher Conrad. *Unsupervised Parameterization for Optimal Segmentation of Agricultural Parcels from Satellite Images in Different Agricultural Landscapes*. MDPI, 2020.

[5] Jonas Schmedtmann and Manuel L. Campagnolo. *Reliable Crop Identification with Satellite Imagery in the Context of Common Agriculture Policy Subsidy Control*. MDPI, 2015.

[6] Thw world bank. Land and food security, 2014. Accessed September 28, 2023. https://www.worldbank.org/en/topic/agriculture/brief/land-and-food-security1/.

[7] United States Geological Survey. What is remote sensing and what is it used for?, 2020. Accessed September 28, 2023. https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used/.

[8] Silvia Liberata Ullo and G. R. Sinha. Advances in iot and smart sensors for remote sensing and agriculture applications. *Remote Sensing*, 13(13), 2021.

[9] European Space Agency. Multispectral instrument (msi) overview, 2023. Accessed September 30, 2023. https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument/.

[10] European Space Agency. Spectral signatures, 2023. Accessed September 30, 2023. https://www.esa.int/SPECIALS/Eduspace_EN/SEMPNQ3Z2OF_2.html/.

[11] GRINDGIS. Vegetation spectral signature cheat sheet, 2017. Accessed September 28, 2023. https://grindgis.com/remote-sensing/vegetation-spectral-signature-cheat-sheet/.

[12] European Space Agency. Sentinel-2 mission guide, 2023. Accessed September 30, 2023. https://sentinel.esa.int/web/sentinel/missions/sentinel-2.

[13] Helder I. Chaminé, Alcides J. S. C. Pereira, Ana C. Teodoro, and José Teixeira. Remote sensing and gis applications in earth and environmental systems sciences. *SN Applied Sciences*, 3(12):870, 2021.

[14] European Space Agency. Sentinel-1, 2023. Accessed September 30, 2023. https://sentinel.esa.int/web/sentinel/missions/sentinel-1.

[15] Carole Planque, Richard Lucas, Suvarna Punalekar, Sebastien Chognard, Clive Hurford, Christopher Owers, Claire Horton, Paul Guest, Stephen King, Sion Williams, and Peter Bunting. National crop mapping using sentinel-1 time series: A knowledge-based descriptive algorithm. *Remote Sensing*, 13(5), 2021.

[16] NASA. What is remote sensing? Accessed September 26, 2023. https://www.earthdata.nasa.gov/learn/backgrounders/remote-sensing/.

[17] NASA. Synthetic aperture radar (sar), 2023. Accessed September 26, 2023. https://www.earthdata.nasa.gov/technology/synthetic-aperture-radar-sar.

[18] Copernicus. Copernicus programme, 2023. Accessed September 26, 2023. https://www.copernicus.eu/en.

[19] European Space Agency. Multispectral instrument, 2023. Accessed September 30, 2023. https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/instrument-payload/.

[20] European Space Agency. Spectral resolution, 2020. Accessed September 26, 2023. https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spectral/.

[21] Vinod Pooja, Ratnadeep Deshmukh, and Priyanka Randive. Vegetation indices for crop management: A review. 01 2018.

[22] Lukas Kondmann, Aysim Toker, Marc Rußwurm, Andrés Camero, Devis Peressutti, Grega Milcinski, Nicolas Longépé, Pierre-Philippe Mathieu, Timothy Davis, Giovanni Marchisio, Laura Leal-Taixé, and Xiao Xiang Zhu1. *DENETHOR: The DynamicEarthNET dataset for Harmonized, inter-Operable, analysis-Ready, daily crop monitoring from space*. MDPI, 2021.

[23] Jesper Rasmussen, Georgios Ntakos, Jon Nielsen, Jesper Svensgaard, Robert N. Poulsen, and Svend Christensen. Are vegetation indices derived from consumer-grade cameras mounted on uavs sufficiently reliable for assessing experimental plots? *European Journal of Agronomy*, 74:75–92, 2016.

[24] Manuel Erena, Salomón Montesinos, Dilia Portillo, Joubert Alvarez, C. Marin, Lara Fernández, J. Henarejos, and L. Ruiz. Configuration and specifications of an unmanned aerial vehicle for precision agriculture. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B1:809–816, 06 2016.

[25] Iqbal H. Sarker. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6):420, 2021.

[26] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[27] Igor M. Coelho, Vitor N. Coelho, Eduardo J. da S. Luz, Luiz S. Ochi, Frederico G. Guimarães, and Eyder Rios. A gpu deep learning metaheuristic based model for time series forecasting. *Applied Energy*, 201:412–418, 2017.

[28] Marc Rußwurm and Marco Körner. Self-attention for raw optical satellite time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169:421–435, 2020.

[29] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[30] Charlotte Pelletier, Geoffrey I. Webb, and François Petitjean. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, 11(5), 2019.

[31] Marc Rußwurm and Marco Körner. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. pages 1496–1504, 2017.

[32] Roberto Interdonato, Dino Ienco, Raffaele Gaetano, and Kenji Ose. Duplo: A dual view point deep learning architecture for time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149:91–104, 2019.

[33] Zhiwei Yi, Li Jia, and Qiting Chen. Crop classification using multi-temporal sentinel-2 data in the shiyang river basin of china. *Remote Sensing*, 12(24), 2020.

[34] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, dec 2017.

[35] M Xu, S Yoon, A Fuentes, J Yang, and D S Park. Style-consistent image translation: A novel data augmentation paradigm to improve plant disease recognition. *Frontiers in plant science*, 12, 2022.

[36] Mingle Xu, Sook Yoon, Alvaro Fuentes, and Dong Sun Park. A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognition*, 137, may 2023.

[37] Suorong Yang, Weikang Xiao, Mengcheng Zhang, Suhan Guo, Jian Zhao, and Furao Shen. Image data augmentation for deep learning: A survey, 2022.

[38] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks, 2015.

[39] Haşim Sak, Andrew Senior, and Françoise Beaufays. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition, 2014.

[40] Abien Fred Agarap. Deep learning using rectified linear units (relu), 2019.

[41] AI4EO. Food security challenge, 2021. Accessed September 28, 2023. https://ai4eo.eu/portfolio/ai4foodsecurity-challenge//.

[42] AI4EO. Food security challenge, 2021. Accessed September 17, 2023. https://ai4eo.eu/portfolio/ai4foodsecurity-challenge/.

[43] International Fetilizer Assoctiation. Fubc: Crop calendars by country.