MSc Artificial Intelligence
Master Thesis

Adapting iTransformer for Early Crop Classification Using Satellite Imagery

by
Guilly Kolkman
11822465

July 7, 2024

48 EC
30-10-2023 to 7-7-2024

Supervisor:
MSc. Cornelis Valk

Examiner:
Dr. Stevan Rudinac

Second reader:
Dr. Arnoud Visser

Universiteit van Amsterdam

Abstract

This research addresses the challenge of early crop classification using satellite imagery, which is important for agricultural monitoring and sustainable practices. In this research the iTransformer architecture, originally designed for time series forecasting, was adapted to a classification model. The iTransformer architecture's inverted embedding was leveraged to capture complex temporal dependencies for crop growth cycles. The utilized dataset consists of Sentinel-1 and Sentinel-2 satellite imagery of the Netherlands with 192.751 parcels and 42 crop classes in 2023. The frequency of observation from Sentinel-2 was inconsistent due to cloud coverage. Nevertheless, the iTransformer's crop classification when using data from the whole year reaches an accuracy of up to 0.96 when using a subset of five classes, and early classification results in an accuracy of 0.89 in June and 0.94 in July. Thus, showing the potential for adapting iTransformer to early crop classification and improving agricultural monitoring.

Acknowledgements

I express my sincere gratitude towards my supervisors, Cornelis Valk and Arnoud Visser. Cornelis Valk, for his guidance and knowledge in the field of remote sensing. Teaching me theoretical and practical knowledge. Arnoud Visser, for his technical knowledge and for aiding in the planning of my thesis. I am thankful to NEO b.v. for giving me the opportunity to perform my master's thesis at their company, using their computational resources, and the great teamwork between colleagues.

Contents

Chapter 1

Introduction

Crop classification is important in agriculture because it aids in agricultural resource management, prediction of yields, and sustainable practices for farming [1]. Sustainable farming practices are important for the United Nations food program and the European Union (EU) Green Deal [2]. One of those sustainable practices is related to the Common Agricultural Policy (CAP). CAP allocates agricultural subsidies throughout the EU, and these subsidies are approximately 50 billion euros per year of the EU's annual budget [3][1]. Monitoring agricultural fields to determine subsidy allocation, is a challenge for the governments. Machine learning-based tools have been developed to automatically classify crops. The Area Monitoring System (AMS) is one of those tools used in Europe. AMS gathers and analyses data on agricultural activities and practices on agricultural parcels using satellites. For example, AMS can identify crop classifications in parcels to determine if farmers are eligible for sustainability subsidies [4]. In addition, it can be used to help identify obligatory adjustments for farmers to become eligible for a subsidy. Currently, farmers apply each year for a subsidy by manually filling in details about their parcels and crops. To reduce this time-consuming step, this research aims to use early crop classification methods. Early crop classification refers to the ability to classify crops before the harvesting season. Achieving accurate early classification can increase efficiency and decision making in agricultural management and policies. This research will focus on crops and parcels of farmers in the Netherlands in the year 2023. Where 1,804 million. ha is the area of cultivated land of which 29,6%, thus 522 thousand ha, is used for agricultural land [5].

Crop classification brings a few challenges with it. There are many different crops, and all crops have a different growth time and a different growth pattern. Grains grow above ground, but asparagus grows underground. Classifying this the data for the entire year is used to account for distinct growth cycles. Another challenge is inconsistent time series data due to cloud coverage. As clouds cover the parcels leading to misinformation.

This research utilizes data from Sentinel-1 and Sentinel-2 satellites from the Copernicus Programme of the EU and the European Space Agency (ESA) in the Netherlands [6]. The satellites have periodic revisit times, 12 days and 5 days respectively. However, due to cloud cover, some satellite imagery can not be used, resulting in inconsistent time series. In addition to the satellite data, the soil type of the various crops is monitored. The soil characteristics, including water retention, influence the crop growth rate. Adding this soil data is expected to improve crop classification accuracy [7].

Various models for crop classification are available, including recurrent neural networks (RNN), long short term memory (LSTM), random forest, Bayesian models, and transformers [8, 9, 10, 11]. This research is based on the iTransformer architecture, a variation of the transformer model introduced in 2017 [12, 13]. Originally iTransformer was designed for time series forecasting. It creates a unique inverted embedding to capture correlations between

[1] `https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview_en`

time series and their multivariate features. In this research, iTransformer was adapted for crop classification, using its strength in temporal dependencies and multivariate feature interactions. Currently, in literature, the iTransformer architecture has not been used for crop classification objectives or for inconsistent time series with various timestamp gaps, as is the case with satellite imagery. Researching time series classification provides insights in the robustness and adaptability of the iTransformer model.

The research questions of this research are:

- Can the iTransformer application be extended from time series forecasting to crop classification?
- To which extent do varying intervals impact the iTransformer model?

By advancing early crop classification methods this research aims to create a methodology with which administrative administrative practices and the efficiency of subsidy allocations for sustainable agricultural practices can be improved.

Figure 1.1: February

Figure 1.2: June

Figure 1.3: October

Figure 1.4: Cloud Coverage

Figure 1.5: Farmland in the province of Friesland over the year

Chapter 2

Theoretical Background

2.1 Related work

Crop classification using satellite imagery has been an active research area in recent years, with various deep learning approaches being explored to improve accuracy and enable early classification. [10, 14, 15, 8, 9, 16]. Zhou et al. conducted a comparative study of 1-DCNN, LSTM, RNN, and random forest architectures using Sentinel-1 data achieving an accuracy of 95-96% for five crop classes in China [17]. Rußwurm et al. used a Sequential Recurrent Encoder with raw Sentinel-2 data reaching an accuracy of 90% across 17 classes in Germany [16]. While these studies focused on whole year classification, recent research in early classification has also been explored. One study introduced an innovative earliness reward loss function that balances early classification with accuracy [9]. It utilizes an LSTM architecture for nine classes, achieving an accuracy of 80% in June. Despite this progress, satellite-based crop classification is unreliable due to cloud coverage, which results in inconsistent time series data. Addressing this issue, Metzger et al. researched the use of Neural Ordinary Differential Equations (NODE) for crop classification under varying cloud cover conditions which demonstrates a novel approach to handling irregularly sampled data in satellite imagery [18]. NODE shows promising results in improving classification accuracy despite inconsistent gaps in the time series data. Satellite imagery is not the only source for crop classification with a vision transformer architecture being utilized for classifying Unmanned Aerial Vehicles (UAV) images [19]. These various approaches highlight the current necessity to enhance crop classification techniques, addressing challenges such as temporal inconsistencies and early crop classification.

2.2 Crops

Crops have growth variations, exhibiting dynamic changes throughout their growth cycles, with each crop displaying unique phenological characteristics [20]. For instance, winter wheat typically grows from January to July and is harvested in July-August. In contrast, Maize grows from April to September, with harvest occurring in October [20]. The growth trajectory of crops is not only influenced by their biological characteristics but is also influenced by environmental and artificial factors. Weather conditions have a crucial role in crop growth, with temperature, precipitation, and sunlight [21]. The type and composition of soil also impact crop growth, with different soil types varying in their capacity to retain moisture after precipitation. Some soils have good water retention while others allow for rapid drainage [22, 23]. Furthermore, the management of farmers introduces additional variability. These practices include their decision on planting dates, irrigation schedules, fertilization, and pest management strategies [24, 25]. This results in crops with the same type displaying differences in their growth patterns through

the growing season. This presents challenges for crop monitoring and classifications with the help of artificial intelligence to account for these dynamic patterns.

2.3 Sentinel-1 and 2

2.3.1 Sentinel-1

Sentinel-1 is a key component of the European Copernicus program [26]. It provides continuous C-band Synthetic Aperture Radar (SAR) observations of Earth in a 12-day repeat cycle with 175 orbits per cycle. Sentinel-1 is made up of a constellation of two satellites, Sentinel-1A and Sentinel-1B, with a 180° orbital phasing difference. Each satellite will fly in a near-polar, sun-synchronized (dawn-dusk) orbit at a height of 693 km [26]. Together they have a 6-day repeat orbit cycle. Currently, there is only one satellite available, Sentinel-1A, due to an electrical failure, which results in a revisit time of 12 days. SAR data is a form of side-looking radar that emits microwave signals and measures the backscattered energy from the Earth's surface. The microwave signal operates at a frequency of 5.405 GHz, which corresponds to a wavelength of about 5.55cm. The advantage of using microwave signals is their capability to penetrate clouds and handle a lack of illumination [1], which results in consistent imaging regardless of any weather conditions. Leveraging the microwave signals is the interferometric wide (IW) swath mode to achieve area coverage. In a single pass, IW covers a 250 km swath with a resolution at 5 x 20 meters, capturing the data in dual polarization. Polarization refers to the orientation of the electromagnetic wave's electrical field of the transmitted signal and how the reflected signal is received. The polarizations this research uses are vertical transmit, vertical receive (VV), and vertical transmit, horizontal receive (VH). These cross-polarized VH and VV backscatter coefficients have a strong ability in monitoring vegetation, and VV-coherence can help detect changes in the crop state [27]. ESA is planning on launching Sentinel-1C and Sentinel-1D to reduce the revisit time, with Sentinel-1C launching in the 4th quarter of 2024 [28].

2.3.2 Sentinel-2

Sentinel-2 is a high-resolution multispectral imaging (MSI) mission, which consists of two satellites, Sentinel-2A, and Sentinel-2B, launched in 2015 and 2017 respectively [29]. Like Sentinel-1 the two satellites are polar orbiting satellites with a revisit time of 5 days or less. The Sentinel-2 mission is mainly dedicated to monitoring agriculture, forests, land-use change, land-cover change, mapping biophysical variables such as the chlorophyll content in leaves, leaf water content, leaf area index as well as risk and disaster mapping [30]. The MSI measures the reflected radiance of 13 spectral bands, from Visible and Near Infrared (VNIR) to Short Wave infra-red (SWIR). This rich spectral information, in combination with the frequent revisit time, makes Sentinel-2 data valuable for agricultural monitoring and crop classification tasks.

The multispectral data from Sentinel-2 offers great use cases for crop studies [27]. The B8 band in particular is useful for assessing the health of the vegetation and chlorophyll content while the near-infrared and shortwave infrared provide insight into the water content and biomass. Combining different spectral bands adds extra vegetation indices to help analyze various crop aspects.

2.3.3 Sentinel-2 Used Spectral Bands

Normalized difference vegetation index (NDVI): This index leverages the contrast between red and near-infrared reflectance to assess the vegetation greenness in the observed area. Values

[1] `https://sentiwiki.copernicus.eu/web/sentiwiki`

Table 2.1: Sentinel-2 bands

Band	Resolution	Central Wavelength	Description
B1	60 m	443 nm	Ultra Blue (Coastal and Aerosol)
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10 m	665 nm	Red
B5	20 m	705 nm	Visible and Near Infrared (VNIR)
B6	20 m	740 nm	Visible and Near Infrared (VNIR)
B7	20 m	783 nm	Visible and Near Infrared (VNIR)
B8	10 m	842 nm	Visible and Near Infrared (VNIR)
B8a	20 m	865 nm	Visible and Near Infrared (VNIR)
B9	60 m	940 nm	Short Wave Infrared (SWIR)
B10	60 m	1375 nm	Short Wave Infrared (SWIR)
B11	20 m	1610 nm	Short Wave Infrared (SWIR)
B12	20 m	2190 nm	Short Wave Infrared (SWIR)

range from -1 to 1 where values below 0 correspond to water, values slightly above 0 indicate bare soil, and 1 means denser and healthier vegetation. The formula for it is (B08 - B04)/ (B08 + B04)

Soil adjusted vegetation index (SAVI): Building upon NDVI, SAVI adjusts for background soil influencing the spectral reflectance. Soil affects the moisture/color/saturation of the red edge vegetation measurements. To correct this, a variable 'L' is added. The value can be between 0 and 1. For this research calculation, the value is 0.5, which is the value for intermediate vegetation densities [31]. The formula for it is (B08 - B04) / (B08 + B04 + L).

Normalized difference turbidity index (NDTI): This index calculates the turbidity or, optical clarity, of water, which indirectly indicates the moisture contents in the soil. A high NDTI means a high clarity of water. The formula is (B04 - B03)/(B04+B03)

Atmospherically resistant vegetation index (ARVI): This index is built upon NDVI, but it is applicable for regions with high atmospheric aerosol content. It includes the blue band and a variable 'y' as a quotient derived from the components of atmospheric reflectance in the blue and red channels. The value 'y' ranges from 0 to 1, with 1 being no atmospheric disturbance. The formula is (B8A - y(2 * B04 - B02)) / (B8A + y(2 * B04 - B02))

Normalized difference infrared index (NDII): This index is a reflectance measurement that is sensitive to changes in the water content of plant canopies. The index values are between -1 and 1 where 1 is a high water content. The formula is (B08 - B11) / (B08 + B11)

Normalized difference index 7 (NDI7): This is similar to NDII, however, it uses a different wavelength, and this wavelength is particularly effective in estimating crop residue cover. The formula is (B08 - B12) / (B08 + B12).

Each of these indices has unique spectral characteristics to provide different information on crop health and associated environmental conditions. By combining these indices a better understanding of crop dynamics can be made, which in turn helps provide a more accurate classification [27].

2.4 Cloud mask

The Sentinel-2 bands are sensitive to clouds, as clouds reflect sunlight, which can interfere with the satellite's ability to accurately capture data [29, 32]. This reflection, especially in the visible

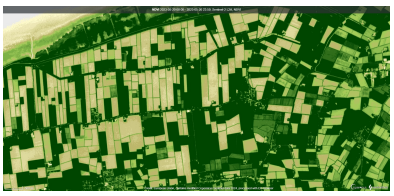

Figure 2.1: NDVI

Figure 2.2: SAVI

Figure 2.3: NDTI

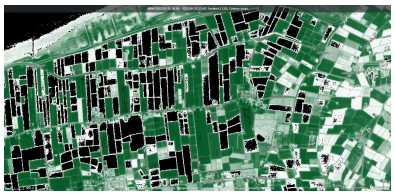

Figure 2.4: ARVI

Figure 2.5: NDII

Figure 2.6: NDI7

Figure 2.7: Different multispectral indices of the same fields as Figure 1.5 in June

and the infrared spectrum, can cause disturbances [33]. A cloud mask is therefore necessary to filter out the data which contain clouds. The cloud mask algorithm is called Sen2Cor. It implements a multi-tiered spectral analysis of the input data. The algorithm first applies a series of thresholds on the reflectance values of Sentinel-2's spectral bands listed in Table 2.1. Additionally, two extra indices, NDVI and the Normalized Difference Snow Index (NDSI), are used for scene classification [34]. Each threshold test generates confidence levels which are combined to generate a probabilistic cloud and snow mask. This resulting mask then undergoes further thresholding to create different cloud probability layers (low, medium, high). Sen2Cor also includes a detection for cirrus clouds, which are thin, wispy clouds. The algorithm utilizes band 2 and band 10 to distinguish high-altitude ice clouds for cirrus threshold classification [35]. The algorithm also detects cloud shadows. These masks help with atmospheric corrections, ensuring more reliable data processing.

2.5 Multilayer Perceptron

A multilayer perceptron (MLP) is a type of neural network characterized by its feed-forward architecture. In deep learning architectures, it is a fundamental building block. It consists of at least three layers of nodes. An input layer, one or more hidden layers, and an output layer. Each neuron, except the neurons from the input layer, computes a weighted sum of its inputs and applies a non-activation function. Activation functions determine whether certain neurons should be activated based on their input. MLPs learn through backpropagation, adjusting the weights between the neurons to minimize the difference between predicted outputs and the actual outputs.

2.6 Transformers

Transformers, first introduced by Vaswani et al [13], were developed to improve on the limitation of recurrent and convolutional neural networks in handling long-term dependencies. They have since had tremendous success in Natural Language Processing (NLP) and computer vision. The model consists of an encoder-decoder structure, each module having a stack of identical layers. The encoder contains multi-head self attention and a fully connected feed-forward network with residual connections followed by layer normalization. The decoder is identical to the encoder except for an added attention layer that attends to the encoder's input.

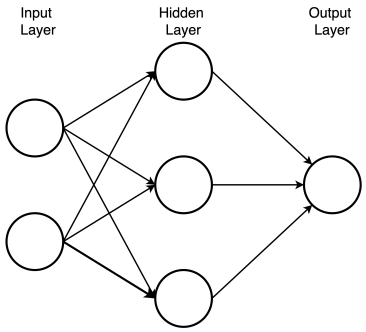

Figure 2.8: Basic MLP architecture

For sequential data like in NLP, embedding layers are used to encode the meaning of input sequences. As transformers process the inputs in parallel, positional information is lost. Therefore, a positional encoding layer is added with a sinusoidal function to indicate which embedding corresponds to which position.

The core innovation of Transformers is the self-attention mechanism which allows the model to dynamically focus on relevant parts of the input sequence. The attention mechanism as originally formulated is:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{2.1}$$

$$softmax : \sigma(x_i) = \frac{exp^{x_i}}{\sum_{j=1}^{n} exp^{x_j}} \tag{2.2}$$

Where Q is a matrix with a set of query vectors, K is a matrix of key vectors, and V is a matrix of value vectors. Q represents the current token which tries to understand its context, with K and V representing the context Q is trying to attend to. In self-attention these three all take the identical input sequence as input. First, a matrix multiplication of Q and K is applied, this step results in attention scores which are then scaled by the square root of the dimension of K. These scores are passed through a softmax function to obtain attention weights. Finally, the attention weights are multiplied by V. This process allows each element to attend to all other elements in the sequence, capturing complex long distance dependencies. Transformers typically use multi-head attention. Multi-head attention extends the basic attention by applying multiple self attention modules in parallel with different learned projections. The output of these multi-heads is then concatenated followed by a linear transformation. This enables the model to capture various aspects of the input simultaneously, making the model well-suited for long distance connections between sequential inputs. The iTransformer has several advantages that make it suitable for crop classification and early classification tasks. Its ability to capture complex temporal dependencies between features due to the inverted embedding approach allows for new insights and flexibility in handling multivariate time series making iTransformer a promising architecture [12]. The following section details how the iTransformer architecture has been adapted to the crop classification problem.

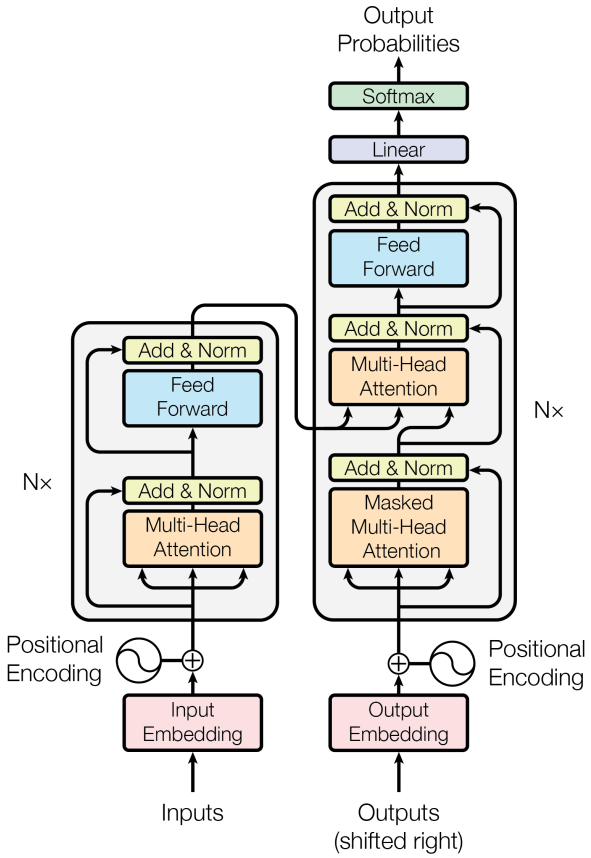

Figure 2.9: Original Transformer architecture with the left block being the encoder and the right block the decoder

Chapter 3

Method

3.1 Dataset

The dataset that was used for this research was provided by NEO b.v. The dataset contains public satellite imagery data from the Copernicus program from 2023, Sentinel-1 and Sentinel-2, from the Netherlands. For both satellites, the data is downloaded from the Copernicus Access Hub. The dataset is preprocessed in a few steps from the original raw satellite images. In the framework, Cirrus detection is implemented to account for the reflection of clouds and snow [29]. The total dataset size for 2023 is 192.751 parcels with 42 classes. Each parcel contains a single time series from 2023. The classes in this dataset are imbalanced with maize more than doubling the second most common class, grains. The top five most common crops are maize with 70.072 data points, grain with 28.279, potato with 22.708, beets with 16.535, and sub grains with 9.125. Labels are provided by the Basisregistratie Gewasparcelen (BRP). [1]

First, a cloudmask filter is used over the Sentinel-2 data to take into account cloud coverage. This is needed because infrared bands reflect on clouds, which results in incorrect values. Secondly, per parcel, the average pixel value is measured on said parcel and this results in each Sentinel-2 band having one value per parcel. This reduces the accuracy on a pixel-level basis but makes the model computationally more efficient. Thus, instead of processing images and pixels, it processes flat values. Daily linear interpolation is used to fill in the missing data points and get data for the entire year. When a time series does not have data due to clouds for example at the start and end of the year, these data points are zero padded.

The features used are VV-coherence, VH-backscatter, and VV backscatter from Sentinel-1. Combination bands, NDVI, SAVI, NDTI, ARVI, NDII and NDI7 from Sentinel-2 and in addition the type of soil for each parcel have been added. Combining Sentinel-1 and Sentinel-2 data improves classification [27]. The soil data were provided by NEO b.v. There are ten types of soil in the dataset namely,

- *Afgesloten zeearmen*: Closed sea arms
- *Duinen*: Dunes
- *Getijdengebied*: Tidal areas
- *Heuvelland*: Hills
- *Hogere Zandgronden*: High sand areas
- *Laagveengebied*: Peatland area

[1] `https://www.pdok.nl/introductie/-/article/basisregistratie-gewaspercelen-brp-`

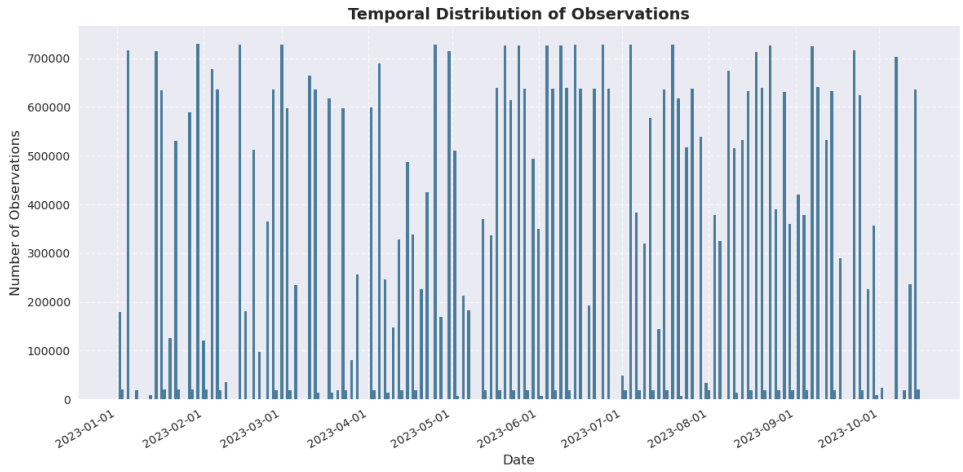

Figure 3.1: Number of observations per date from Sentinel-2 data

- *Noordzee*: Northsea
- *Rivierengebied*: River area
- *Zeekleigebied*: Marine clay area
- *Niet indeelbaar*: No category

This results in 20 features in total.

The dataset was split into training, validation, and test sets in a ratio of 70:20:10. The dataset crop labels are provided by farmers, which makes the labels prone to human error. The dataset is also inconsistent due to the cloud coverage, with some dates having more than 700.000 observations and some dates having less than 8000 observations.

3.1.1 iTransformer

The original iTransformer proposed by Liu Yong [12] for multivariate time series forecasting tasks uses the encoder only architecture from the basic transformer. This includes the embedding layer, projection, and transformer blocks.

3.1.2 Base iTransformer multivariate forecasting

Given historical observations $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_T\} \in \mathbb{R}^{T \times N}$ with $\mathbf{T}$ time steps and $\mathbf{N}$ number of variates, the future $\mathbf{S}$ timesteps $\mathbf{Y} = \{\mathbf{x}_{T+1}, \ldots, \mathbf{x}_{T+S}\} \in \mathbb{R}^{S \times N}$ are predicted in multivariate time series forecasting.

iTransformer changes the basic architecture by inverting the embedding layer, unlike traditional-based time series forecasting which treats multiple variables at a single time point as a token. iTransformer embeds the whole time series of each variable as a single token, which results in the model being able to capture the intrinsic properties of each of the variable's time series more effectively.

$$h_n^0 = Embedding(X_{:,n}) \tag{3.1}$$

$$H^{l+1} = TrmBlock(H^l), l = 0, ..., L-1 \tag{3.2}$$

$$\hat{\mathbf{Y}}_{:,n} = Projection(h_n^L) \tag{3.3}$$

Here, $\mathbf{H} = \{h_1, ..., h_N\} \in \mathbb{R}^{NxD}$ where N is the embedded tokens of dimension D and the superscript denotes the index layer. Both the embedding and projection are implemented by a

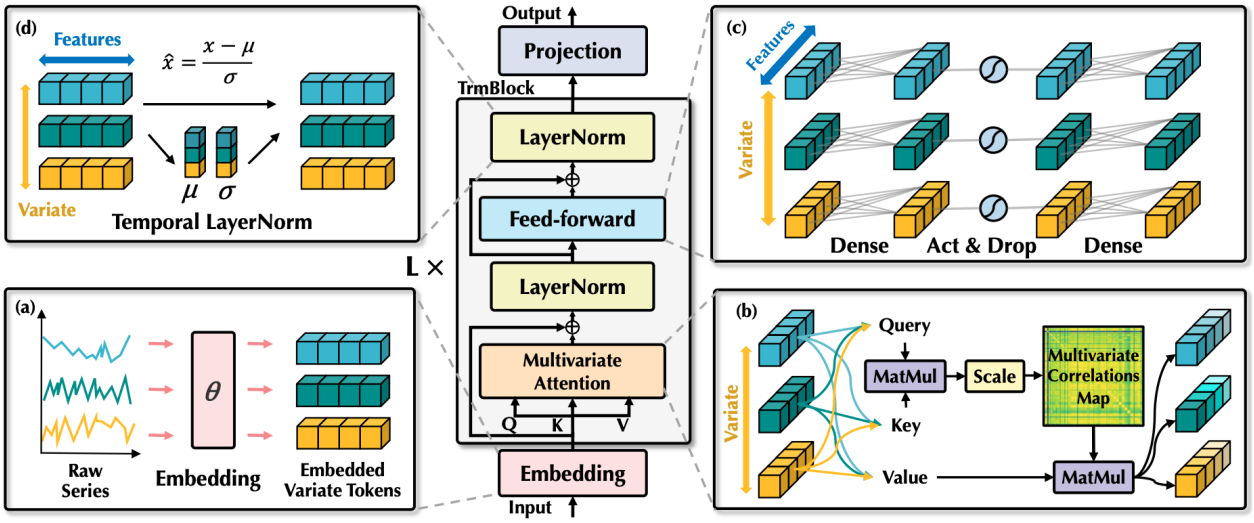

Figure 3.2: iTransformer architecture. Courtesy [12]

multi-layer perceptron (MLP). The resulting embedded variate tokens interact through the self-attention mechanism. Which is then independently processed by a shared feed forward network within each transformer block. Moreover, with this approach, the sequential order of the data is implicitly stored in the neurons of the feed-forward network. The added benefit is that there is no need for the position embedding from the vanilla transformer [12]. The iTransformer consists of L blocks with each containing layer normalization, a feed-forward network, and a self attention module.

Layer normalization was originally proposed to enhance the convergence and training stability of deep learning networks [36]. Where in normal Transformers, layer normalization normalizes the multivariate representation at each timestamp. Instead, iTransformer applies layer normalization the the series representation of each variable. This method has been demonstrated to be effective in addressing non-stationary issues in time series analysis. By normalizing each time series to a Gaussian distribution, discrepancies arising from inconsistent measurements can be mitigated in different variables [37] [38]. Originally in Transformers normalization was applied across different time steps, resulting in over-smoothing of the time series. iTransformer preserves the temporal dynamics situated with each variable.

$$\text{LayerNorm}(\mathbf{H}) = \left\{ \frac{\mathbf{h}_n - \text{Mean}\left(\mathbf{h}_n\right)}{\sqrt{\text{Var}\left(\mathbf{h}_n\right)}} \middle| \; n = 1, \ldots, N \right\} \tag{3.4}$$

3.1.3 Classification iTransformer

The main architecture of iTransformer stays unchanged since the strength of the iTransformer architecture is the learned representations of features with time series. The projection layer returns to the number of features present in the dataset. From these features a linear layer going from the number of features to the number of classes with a softmax activation function is used to obtain the probabilities of each class. The loss is changed from MSE to cross-entropy loss with label smoothing and class weights. Cross-entropy is defined by the Pytorch loss function of cross-entropy which equals.

$$l(x, y) = -\sum_{c=1}^{C} w_c log \frac{exp(x_c)}{\sum_{i=1}^{C} exp(x_i)} y'_c \tag{3.5}$$

Where x is the input logits from the model and y' is the smoothed target distribution. For the correct class, this equals $1 - \alpha$ and for other classes, it is $\alpha/C - 1$ where $\alpha = 0.1$ is the smoothing parameter. w_c is the weight for class c and C is the number of classes. This research uses label smoothing because the ground truth was given by local farmers, which causes some discrepancies to arise. Class weights are introduced to handle the imbalanced class dataset. These weights give more importance to classes that occur less in the dataset. This implementation of cross-entropy loss allows the model to be more robust due to label smoothing, to handle unbalanced classification better, and to improve probability estimates. The optimizer is the Adam optimizer.

3.2 Hyperparameters

The iTransformer model was configured with the following parameters.

Table 3.1: Hyperparameters for the iTransformer model

Hyperparameter	**Value**	**Description**
Learning rate	0.00001	Initial learning rate
Batch size	32	Number of samples in a training loop
Epochs	25	Number of passes through the dataset
Patience	4	Number of epochs with no improvement before stopping
Layer size	2048	Dimensionality of the neural network
Attention heads	4	Number of multi-head attention modules
Encoder layers	4	Number of encoder layers
Dropout rate	0.1	Probability of dropping out a neuron during training
Optimizer	Adam	Optimization algorithm used for updating the weights
Starting month	March	First month of data used

Chapter 4

Experiments

4.1 Metrics

To measure model performance on the multi label classification task this research used a few commonly used metrics: overall accuracy, recall, precision, and the F1-score. Accuracy tells us the overall correctness of the model across all classes. It calculates the ratio of correctly predicted against the total number of instances. Precision calculates the relevance of the retrieved results. It shows when predicting the target class how often these predictions are correct. Recall, also known as sensitivity, calculates the relevant instances retrieved against the total number of relevant instances in the dataset. It shows out of all the positives how many are correctly classified. The F1 score is the harmonic mean between precision and recall. This score is useful when there is an imbalance in the classes like the NEO dataset that is used in this research. Accuracy over time is also calculated. Here for each timestep in the dataset accuracy is calculated and those values stored, with the expectation that over time the accuracy will improve The formulas are given by: [4.1, 4.2, 4.3, 4.4. Where TP is True Positives, TN is True Negatives, FP is False Positives and FN is False Negatives.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{4.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{4.4}$$

4.2 Whole year classification

Using satellite data has the disadvantage that the time series are based on the visiting rate of the satellites. This results in inconsistent time series, meaning the experiments in this research are based on this inconsistent time interval. Two main experiments were performed: One for time intervals from 5-12 days with the whole year as input data, so the data are from January till December and the other for early crop classification. The intervals were chosen based on the Sentinel satellite's revisit time, with Sentinel-2 having a revisit time of 5 and Sentinel-1 having a revisit time of 12 days. Five classes were chosen as a baseline: grains, onion, chicory, peony, and turnips. The choice of classes was based on their different growth cycles, growth characteristics, and the number of data points. A 5-day interval has the best scores over all

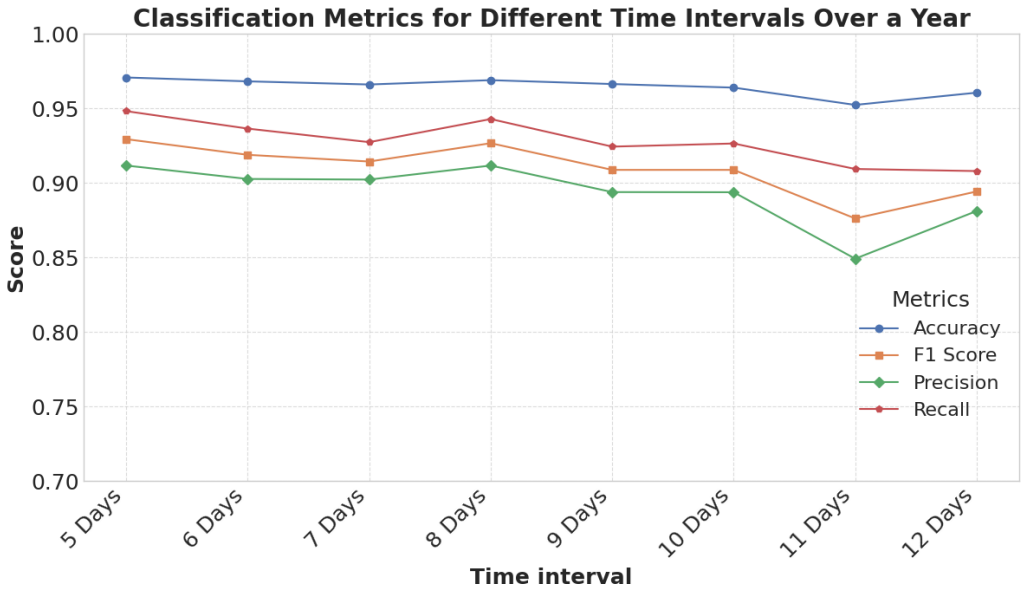

Figure 4.1: Classification of crops, using the entire year (March-December) as training data

metrics with a steady decline the longer the interval is. The first two months, January and February are skipped when loading the data because in these months no crops are planted in the Netherlands. An experiment was performed to test this, with parameters from a 5-day interval. This resulted in an accuracy of 0.13, precision of 0.54, recall of 0.78, and F1 score of 0.48.

4.3 Early classification

As can be seen in figure 4.1 the 5-day interval demonstrated the highest accuracy among the tested intervals. Therefore, the parameters used for this interval are used for early classification. Each month individual models were trained, where the training data was capped at the end of each respective month. For instance, the training data for July contains satellite data from March through the end of June.

Training multiple models has the advantage of learning seasonal variations with the addition of one month for every model. It also supports consistency in the training data because the model inputs have identical sequence lengths without too much zero padding.

4.4 Number of classes

This research uses five classes as a proof of concept. However, the full dataset contains 42 classes in total. When training the model on 42 classes, the overall performance metrics significantly decline. Training on, again, the parameters from the best whole year classification, the scores result in the following accuracy: 0.32, Precision: 0.46, Recall: 0.53, and F1: 0.44. In Figure 4.4 *paksoi* is not represented in the True labels because of a small number of samples. Which resulted in no entries of *paksoi* in the test set. Moreover, there is one class with a lot of false predictions, which is also *paksoi*. This shows that the model has trouble recognizing crops with almost no entries in the dataset, due to class imbalance.

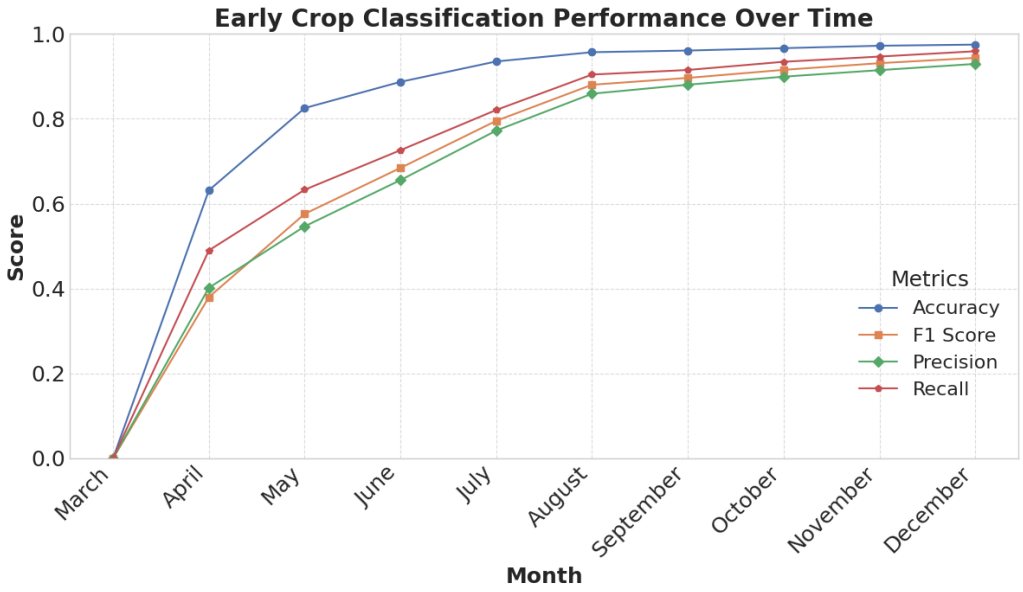

Figure 4.2: Early classification

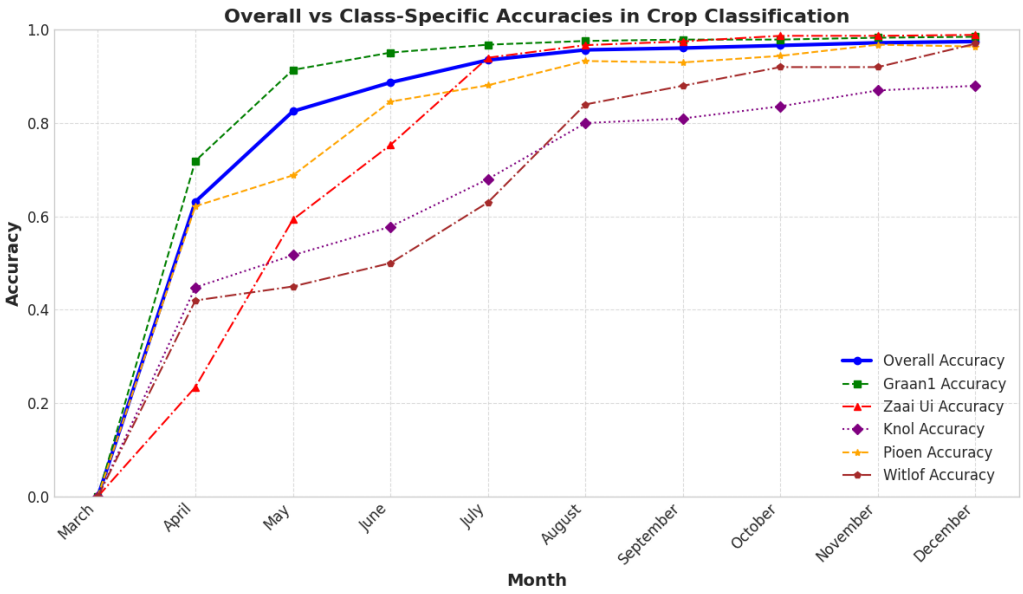

Figure 4.3: Accuracy of 5 used classes and overall accuracy

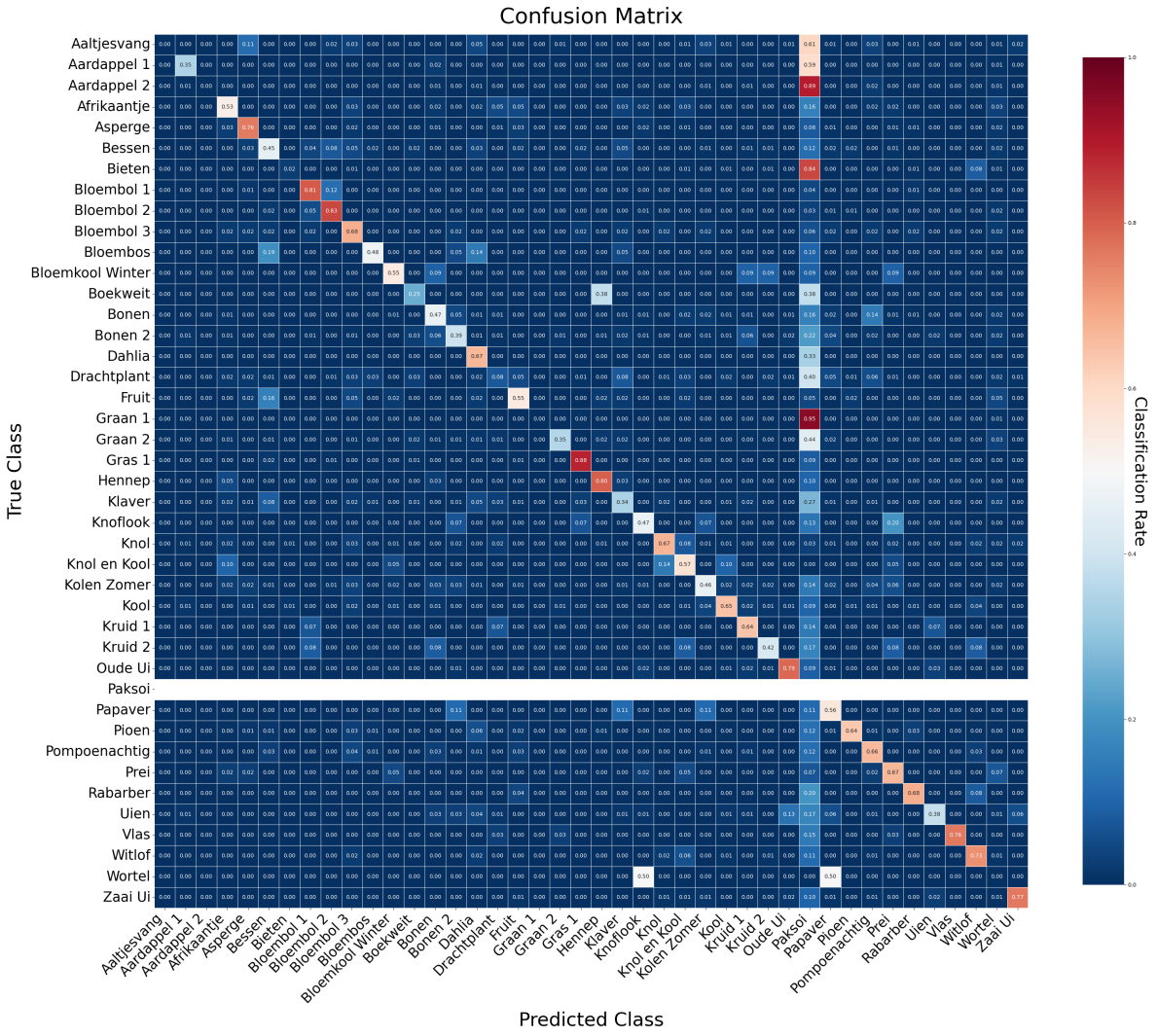

Figure 4.4: Confusion matrix of the 42 classes

4.4.1 Imbalanced classes

To further investigate the impact of class imbalance, another experiment was performed using maize, the most represented crop in the dataset. Maize has 70.072 entries in the dataset more than doubling the next most represented class, grains with 28,279 entries. With maize having a classification accuracy of 0.96, precision of 0.82, recall of 0.93, F1 of 0.86.

4.5 Hyperparameter Tuning

Table 4.1: Hyperparameter testing for the iTransformer model

Hyperparameter	**Tested**	**Best**
Learning rate	0.1, 0.01, 0.001, 0.0001, 0.00001	0.00001
Batch size	32	-
Patience	3, 4, 5	4
Layer size	512, 1024, 2048	2048
Attention heads	2, 4, 8	4
Encoder layers	2, 4 ,7	4
Dropout rate	0.1	-
Optimizer	Adam	-
Starting month	January, February, March	March

Chapter 5

Discussion

The experiments in this research reveal several important insights into crop classification using the iTransformer model and satellite imagery. The results demonstrate that shorter time intervals between satellite observations lead to better performance across all metrics. 5-day intervals perform the best in comparison to other time intervals. This could be caused by the model learning to capture subtle changes in the crop growth patterns, resulting in a more accurate classification. It also shows that the standard 5-day interval from Sentinel-2 revisit time is suitable for crop classification, with a slight decrease in performance the larger the time interval. However, a 12-day interval has fewer data points in a year which results in less complexity. This trade-off of a decrease in accuracy or less computationally expensive training is to be considered.

When examining the tests for individual classes, it is apparent that different crops have varying optimal classification periods, with some crops, like chicory getting a jump in accuracy between July and August. This variation can be attributed to the different growth cycles and the phenological characteristics of each crop. For example, some crops are better identifiable during their sprouting or flowering stages, while others are identified during sowing.

Scaling from 5 to 42 classes noticed a considerable decrease in model performance. This indicates the difficulty in the multi-class classification of crops with varying growth and crop characteristics. Some crops, during growth, can exhibit similar spectral and temporal footprints. However, as shown in the experiment involving maize. Adding a well-represented class does not decrease the performance metrics by a large margin. This indicates that crops with a significant amount of data points can be accurately classified. Another solution would be to develop a hierarchical classification approach to handle the classification of lesser observed crops, and aggregating similar species of crops and refining the classification within these groups. This potentially reduces the impact of class imbalance.

An important finding from the experiments is the inclusion of January and February in the classification process. This can be attributed to the lack of agricultural activity during winter. Most parcel data exhibit similar characteristics since fields are barren around that time, which leads to confusion in the models' training. This result makes clear that selecting the appropriate periods for training and classification leads to major differences.

A practical application of this research is the potential to verify farmer-provided crop information. In the Netherlands, farmers register their crop types in May as part of the agricultural subsidy process. Therefore, the results of our early classification could provide a useful tool for cross-checking the farmer-submitted data, by reducing the need for on-site inspections and helping identify discrepancies and errors in crop registration.

The challenge of inconsistent time series due to cloud coverage remains an issue in satellite-based crop classification. In this research linear interpolation was used to account for cloud coverage, however, a more complex method could be used for interpolation, such as Neural

Ordinary Differential Equations. Researching a solution for cloud coverage can be crucial, especially in regions like the Netherlands, where cloud frequency increases towards the end of the year.

Adapting the iTransformer architecture from time series forecasting to classification problems shows promise. Whole year and early classification tasks show that the architecture can learn crop growth patterns between individual crops and seasons with a few differentiable crops. However, the model's performance decreases, and handling a large number of classes suggests areas for further improvement.

5.1 Research Journey and Lessons Learned

Throughout this research, several challenges were encountered. This resulted in changes to the methodology and structure of the research and provided valuable learning experiences. Initially, this research aimed to utilize the NODE model architecture, with a focus on solving inconsistent time series for satellite imagery data. However, the implementation of the NODE model presented difficulties. The original GitHub repository, which was used as a starting point, described a dataset with characteristics that differed from our dataset. This resulted in significant code compatibility issues when applying NODE to the dataset of this research. Proving more time-consuming and complex than originally anticipated. After a month and a week of attempts to solve these compatibility issues with no success, it became clear that a different approach was needed.

This is where the focus shifted to the iTransformer architecture. This proved to be fruitful in terms of the research progress. However, this late shift in also resulted in time constraints for the research. The reduced time frame limited the research capability to conduct a comprehensive cross-validation, which could affect the robustness and generalization of the results. This experience taught me that under time constraints, efficient decision-making needs to be made when facing obstacles.

Computational resource limitations presented another challenge. Although GPU resources were available in NEO, these resources were shared amongst the entire company. Sharing these resources sometimes resulted in queuing times for running tests and experiments.

The data prepossessing also gave me an additional challenge of dealing with memory management. The process of combining multiple large datasets from Sentinel-1 and Sentinel-2 matching them on parcels and dates, proved to be memory intensive. This resulted in a shortage of RAM and the program crashing. Therefore, I explored new implementations and optimization techniques to handle large amounts of data.

In conclusion, while these challenges affected the research, they served me for significant personal and professional growth, in particular when considering practical experiences in problem-solving, adaptability, and various research methodologies.

Chapter 6

Future Work

Future research in crop classification using remote sensing could be built upon in several directions. One key direction is the development of a hierarchical classification approach to address the imbalanced datasets. This could involve aggregating crops with similar phenological features into sub-classes, followed by a classification within these sub-groups.

Another area for improvement is the integration of additional data sources. Incorporating data such as meteorological data, could provide more context for crop growth patterns in the model.

Expanding the dataset to include multiple years of data could account for cross-season variability, potentially improving the model's robustness. This would help to distinguish seasonal fluctuations and long term trends in crops. These additions could help improve crop classification accuracy and generalizability of the model. Further improving crop monitoring.

Chapter 7

Conclusions

In this research, two research questions were studied. The potential to adapt iTransformer to a crop classification task and the impact of varying time intervals on the model's performance. For the first question, this research demonstrates that extending the iTransformer architecture from time series forecasting to crop classification is possible, showing promising results for whole-year classification and early crop classification when dealing with a select amount of classes. The iTransformer's ability to learn temporal dependencies and multivariate interactions proves useful when distinguishing between various crop types. However, increasing the amount of classes and having an imbalanced dataset decreases the model's performance significantly. This suggests that further research may be needed to handle the complexities of real-world multi-class crop classifications. When considering the second research question, this research's experiments demonstrate that varying time intervals impact the iTransformer model's performance. Shorter time intervals of Sentinel-2's 5-day interval perform better than their Sentinel-1's 12-day counterpart. This yields better results overall metrics by a slight margin. While challenges remain, this research demonstrated the potential of the iTransformer model for crop classification using satellite imagery. Expanding the field of crop monitoring with a new architecture for crop classification.

Appendix A

Optional appendix

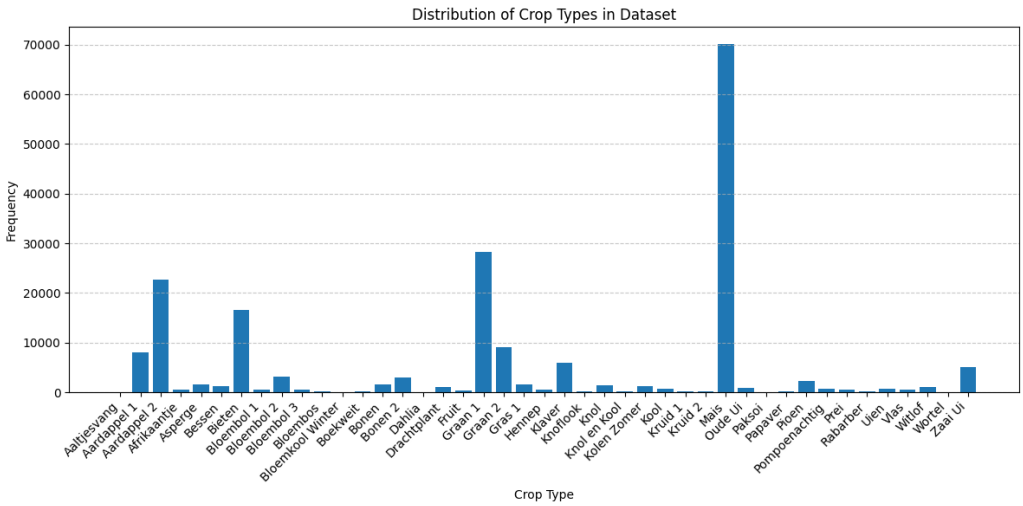

Figure A.1: Crop distribution in the dataset

Bibliography

[1] D.K. Sreekantha and Kavya A.M. *"Agricultural crop monitoring using IOT - a study"*. In: (2017), pp. 134–139. DOI: 10.1109/ISCO.2017.7855968.

[2] Alexandros Tataridas et al. *"Sustainable crop and weed management in the era of the EU Green Deal: A survival guide"*. In: *Agronomy* 12.3 (2022), p. 589.

[3] Félix Quinton and Loic Landrieu. *"Crop rotation modeling for deep learning-based parcel classification from satellite time series"*. In: *Remote Sensing* 13.22 (2021), p. 4599.

[4] Jonas Schmedtmann and Manuel L Campagnolo. *"Reliable crop identification with satellite imagery in the context of common agriculture policy subsidy control"*. In: *Remote Sensing* 7.7 (2015), pp. 9325–9346.

[5] P. Ramaekers P. Berkhout H. van der Meulen. *Staat van Landbouw, Natuur en Voedsel*. Tech. rep. CBS, 2023.

[6] Zbyněk Malenovský et al. *"Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land"*. In: *Remote Sensing of Environment* 120 (2012), pp. 91–101.

[7] JF Bierhuizen and NM De Vos. *The effect of soil moisture on the growth and yield of vegetable crops*. Tech. rep. Institute for Land and Water Management Research, 1959.

[8] John Brandt. *"Spatio-temporal crop classification of low-resolution satellite imagery with capsule layers and distributed attention"*. In: (2019). eprint: arXiv:1904.10130.

[9] Marc Rußwurm et al. *"End-to-end learned early classification of time series for in-season crop type mapping"*. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 196 (2023), pp. 445–456.

[10] Yongchuang Wu et al. *"Remote Sensing Crop Recognition by Coupling Phenological Features and Off-Center Bayesian Deep Learning"*. In: *MDPI Remote sensing* (2023).

[11] Hongwei Zhao et al. *"Evaluation of Three Deep Learning Models for Early Crop Classification Using Sentinel-1A Imagery Time Series-A Case Study in Zhanjiang, China"*. In: *Remote Sensing* 11 (Nov. 2019), p. 2673. DOI: 10.3390/rs11222673.

[12] Yong Liu et al. *"itransformer: Inverted transformers are effective for time series forecasting"*. In: *arXiv preprint arXiv:2310.06625* (2023).

[13] Ashish Vaswani et al. *"Attention is all you need"*. In: *Advances in neural information processing systems* 30 (2017).

[14] Yongchuang Wu et al. *"Remote Sensing Crop Recognition by Coupling Phenological Features and Off-Center Bayesian Deep Learning"*. In: *Remote Sensing* 15.3 (2023).

[15] Ayshah Chan, Maja Schneider, and Marco Körner. *"XAI for Early Crop Classification"*. In: *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*. 2023, pp. 2657–2660.

[16] Marc Rußwurm and Marco Körner. *"Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders"*. In: *ISPRS International Journal of Geo-Information* 7.4 (2018).

[17] Hongwei Zhao et al. *"Evaluation of Three Deep Learning Models for Early Crop Classification Using Sentinel-1A Imagery Time Series—A Case Study in Zhanjiang, China"*. In: *Remote Sensing* 11.22 (2019).

[18] Nando Metzger et al. *"Crop Classification Under Varying Cloud Cover With Neural Ordinary Differential Equations"*. In: *IEEE Transactions on Geoscience and Remote Sensing* PP (Aug. 2021), pp. 1–12.

[19] Reenul Reedha et al. *"Transformer Neural Network for Weed and Crop Classification of High Resolution UAV Images"*. In: *Remote Sensing* 14.3 (2022).

[20] Liheng Zhong, Lina Hu, and Hang Zhou. *"Deep learning based multi-temporal crop classification"*. In: *Remote Sensing of Environment* 221 (Dec. 2018), pp. 430–443.

[21] Porter John R and Semenov Mikhail A. *"Crop responses to climatic variation"*. In: *Phil. Trans. R.Soc* 360 (2005), pp. 2021–2035.

[22] Douglas L Karlen, NS Eash, and PW Unger. *"Soil and crop management effects on soil quality indicators"*. In: *American Journal of Alternative Agriculture* 7.1-2 (1992), pp. 48–55.

[23] RD Connolly. *"Modelling effects of soil structure on the water balance of soil–crop systems: a review"*. In: *Soil and Tillage Research* 48.1-2 (1998), pp. 1–19.

[24] Verónica Saiz-Rubio and Francisco Rovira-Más. *"From Smart Farming towards Agriculture 5.0: A Review on Crop Data Management"*. In: *Agronomy* 10.2 (2020).

[25] David J Connor, Robert S Loomis, and Kenneth G Cassman. *Crop ecology: productivity and management in agricultural systems*. Cambridge University Press, 2011.

[26] Pierre Potin et al. *"Sentinel-1 Mission Status"*. In: *Procedia Computer Science* 100 (2016), pp. 1297–1304.

[27] Amal Chakhar et al. *"Improving the Accuracy of Multiple Algorithms for Crop Classification by Integrating Sentinel-1 Observations with Sentinel-2 Data"*. In: *Remote Sensing* 13.2 (2021).

[28] Andreas Braun. *"Retrieval of digital elevation models from Sentinel-1 radar data–open applications, techniques, and limitations"*. In: *Open Geosciences* 13.1 (2021), pp. 532–569.

[29] Magdalena Main-Knorn et al. *"Sen2Cor for Sentinel-2"*. In: *Image and Signal Processing for Remote Sensing XXIII*. Ed. by Lorenzo Bruzzone. Vol. 10427. International Society for Optics and Photonics. SPIE, 2017, p. 1042704.

[30] ESA. *"Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services (ESA SP-1322/2 March 2012)"*. In: (2012).

[31] Alfredo Huete. *"Huete, A. R. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment"*. In: *Remote Sensing of Environment* 25 (Aug. 1988), pp. 295–309.

[32] R. Richter, J. Louis, and B Berthelot. *Sentinel-2 MSI – Level 2A Products Algorithm Theoretical Basis Document. ESA Report, ref S2PAD-ATBD-0001*. Tech. rep. ESA, 2011.

[33] Louis Baetens, Camille Desjardins, and Olivier Hagolle. *"Validation of Copernicus Sentinel-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure"*. In: *Remote Sensing* 11.4 (2019).

[34] Sergii Skakun et al. *"Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2"*. In: *Remote Sensing of Environment* 274 (2022), p. 112990.

[35] Dirk Tiede et al. *"Investigating ESA Sentinel-2 products' systematic cloud cover overestimation in very high altitude areas"*. In: *Remote Sensing of Environment* 252 (2021), p. 112163.

[36] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. *"Layer normalization"*. In: *arXiv preprint arXiv:1607.06450* (2016).

[37] Taesung Kim et al. *"Reversible instance normalization for accurate time-series forecasting against distribution shift"*. In: *International Conference on Learning Representations*. 2021.

[38] Yong Liu et al. *"Non-stationary transformers: Exploring the stationarity in time series forecasting"*. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 9881–9893.