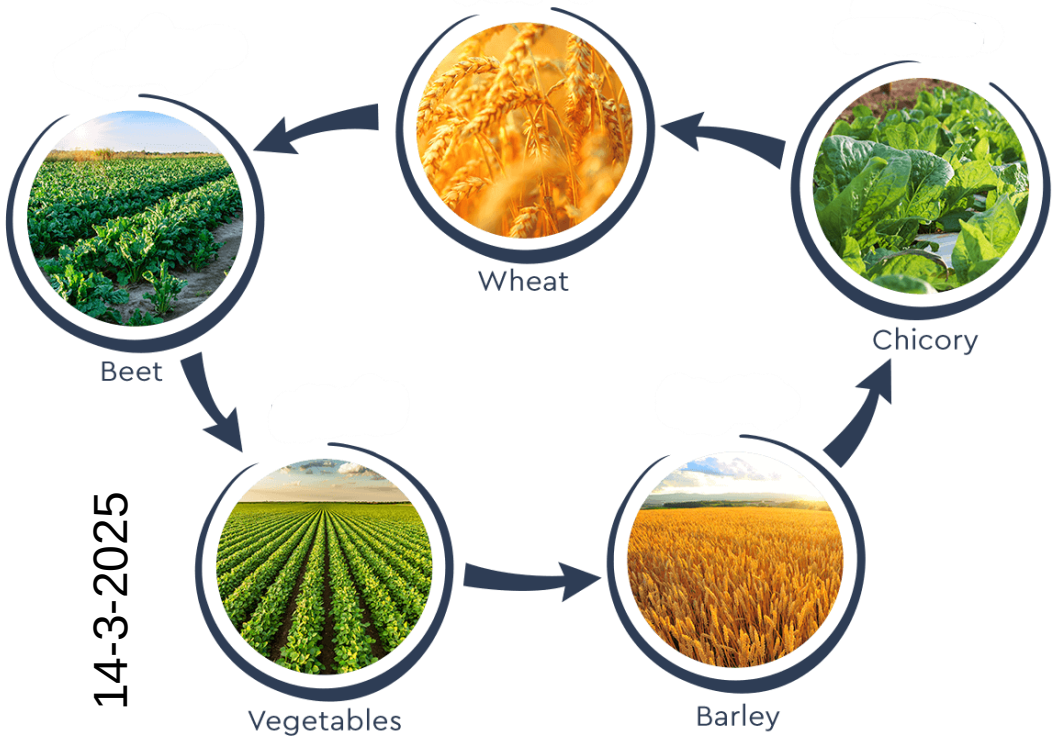
Geo-information Science and Remote Sensing

Thesis Report GIRS-2025-12

Leveraging data science and machine learning to analyse and predict main crop rotation patterns

Wessel Eshuis



WAGENINGEN
UNIVERSITY & RESEARCH

Leveraging data science and machine learning to analyse and predict main crop rotation patterns

Wessel Eshuis

Registration number 1049216

Supervisors:

Marc Rußwurm

Paolo Dal Lago

A thesis submitted in partial fulfilment of the degree of Master of Science
at Wageningen University and Research,
The Netherlands.

14 March 2025

Wageningen, The Netherlands

Thesis code number: GRS-80406
Thesis Report: GIRS-2025 -12
Wageningen University and Research
Laboratory of Geo-Information Science and Remote Sensing

Abstract

Crop rotations play a fundamental role in effective agricultural management. They enhance soil fertility, reduce pests and diseases and ultimately improve crop yield. This thesis is a deep dive into seven-year crop rotation patterns across Dutch agricultural fields, utilizing data from the national field parcel dataset. The thesis is split into two distinct approaches. The first objective is to recognize, characterize and spatially map crop rotation patterns at both a national and regional scale. The second objective utilizes this knowledge as a foundation for evaluating the performance of a newly developed transformer-encoder model, which has been trained to predict future crops using the national field parcel dataset. The model is intended to form the basis of a future AI-driven decision support system, enabling the simulation of the impact that policy change could have on agricultural practices, such as the influence of nitrogen regulations on the use of cover crops in farming systems.

18 major distinct crop rotation patterns were revealed using hierarchical clustering. This clustering method was done using a Hamming distance matrix as input, which quantified similarity between all sequences, allowing the clustering process to reveal distinct patterns in crop rotation practices. Through spatial analysis, the identified major crop rotation clusters were examined in relation to key variables such as: soil texture, crop diversity and the use of cover crops. Providing insights into the underlying factors influencing crop rotation patterns as well as the effect of these patterns on the agricultural landscape.

Upon validating the crop rotation patterns, a transformer-encoder model was developed. This deep learning approach has been trained to learn and recognize crop rotation patterns, enabling the prediction of likely crop choices for the following season based solely on previously cultivated crops. The model achieves a top-3 prediction accuracy exceeding 80%. Its performance was evaluated against three, non-machine learning predicting approaches, by using the Kullback-Leibler divergence. Measuring the divergence from the reference distribution and predicted distributions. The model outperformed two of these approaches and demonstrated the potential to surpass the third. The results show that the model can predict across varying sequence lengths and rare crop sequences, highlighting its robustness compared to the alternative predictive methods.

Keywords: *Crop rotation, Hamming distance, Hierarchical clustering, KL-divergence Transformer-encoder, AI-driven decision support system*

Contents

1 Introduction

Effective implementation of agricultural policies plays a critical role in the balancing act that farmers face between food production, environmental sustainability and economic viability (Lankoski & Thiem, 2020). However, implementing such policies is far from straightforward. An intriguing example of this is the implementation of cover crops in the Dutch farming system, which has been driven by the European Union's push to improve water quality across Europe.

Cover crops or "catch crops", are plants grown between main crop cycles. Next to improving soil health and structure, preventing erosion, increasing the amount of organic matter and their role as green manure crops, they play a significant role in role in tackling environmental issues, particularly in reducing nitrogen leaching (Kaspar & Singer, 2011). Nitrogen leaching occurs when excess nitrogen is washed away by rainfall into the local water system. Which happens when there is no main crop on the field that can take up this nitrogen. This leads to water pollution, harming aquatic ecosystems and human water supplies (Liu et al., 2024). Cover crops retain this excess nitrogen in their roots and biomass, effectively "catching" the nitrogen before it is lost. The European Nitrates Directive (European Union, 1991) was adopted in the nineties to reduce water pollution caused by nitrogen leaching from agricultural practices. Setting a threshold at 50 mg/L of nitrate in surface waters. In the Netherlands first Nitrates Action Programme regulated manure management and promoted crop rotation, as well as encouraging the use of cover crops to reduce nitrogen leaching. These measures were revised in 2004 and led to the obligation to plant a cover crop after the cultivation of maize on sandy soils (Ministerie van Landbouw, Natuur en Voedselkwaliteit, 2005). After the review of the implementation of these directives in 2010, the European commission concluded that some regions were still facing high nitrate levels in ground- and surface water. Resulting in the adaption of a new Dutch nitrate plan in 2014 (Ministerie van Landbouw, Natuur en Voedselkwaliteit, 2014), with cover crops becoming one of the central features. Farmers could get more subsidy through their sustainability score if they planted cover crops, with a specific focus on regions that appeared to be prone to nitrogen leaching. In 2015 the European court ruled that the Netherlands did not comply with the nitrate directive. The court required the Dutch government to take immediate action to reduce the nitrate levels in the water systems, this resulted in the 6th Nitrate Action Programme (Ministerie van Landbouw, Natuur en Voedselkwaliteit, 2017). Mandating the use of covers by October 1st in nitrogen-leaching sensitive areas. Non-compliance with this planting deadline resulted in a reduction in the amount of manure allowed per hectare, providing a significant financial incentive for farmers to adopt cover cropping practices. In 2022, these regulations became even stricter, with further reduction in allowable manure application for failure to implement the required cover crops, as the Netherlands continued to exceed the to meet the 50 mg/L concentration limit (Ministerie van Landbouw, Natuur en Voedselkwaliteit, 2021).

This example highlights the importance of properly implementing agriculture policies. Predicting the effects of policy changes could lead to more effective policymaking, helping to avoid the frequent alteration of regulations, as happened with the nitrogen case. However, predicting policy outcomes is challenging due to the dynamic nature of agricultural systems, where farmer decision-making processes are far from uniform. And are dependent on a wide range of variables. Without accurate modelling, forecasting, and thorough analysis of existing policies, policymakers risk implementing measures that fail to achieve their intended outcomes.

1.1 Data science

As the availability of data on farming farm practices continues to grow, large-scale data analysis becomes increasingly feasible. In agriculture, and especially in the Netherlands, the government gathers field parcel data covering the history of every field in the country. These datasets offer opportunities to apply data analysis techniques to better understand current farming practices, such as crop rotation and cover crop use. By tracking the cultivation per field over time and across regions, the identification of trends and patterns becomes possible. Which in turn can be used to explore correlations with policy changes.

In the context of crop rotation analysis, Stein and Steinmann (2018) propose a crop sequence typology that emphasizes both functional and structural diversity within crop rotations. Such a sequence typology can be applied to spatial data to identify regional trends in space as well as over time, as showed by Ballot et al. (2023). After simplifying the crop sequences by considering crop groups like cereals, corn or root crops instead of individual species, they managed to identify 8 different crop rotations over space and time. This was done using a hierarchical clustering method as a data science approach, they found 8 clear crop rotations and their distribution on a European scale.

In a different field of science, sequences are compared using edit distances. These distances are computed by counting the minimum amount of changes need to go from one sequence to the other. Like for example the soft edit distance for genetic sequence analysis, which makes it easier to cluster using variable-length sequences (Ofitserov et al., 2019). Or the Hamming distance (Hamming, 1950), which counts the number of substitutions between two sequences. For example, this method is used in comparing DNA sequences (Al Kindhi et al., 2017).

Referring back to the crop sequence typology of Stein and Steinmann, it would be a novelty to use language processing techniques to compare cropping sequences. It would also be an interesting to use such an edit distance matrix as input in the hierarchical clustering of crop rotation patterns, as this is a data science approach yet to be explored.

1.2 Machine learning

When analysing extensive datasets in data science, machine methods should nowadays be considered due to their ability to recognize and learn complex patterns on a large scale. These models, and in particular the deep learning models, have the potential to identify underlying relationships between variables, enabling predictive analytic and uncovering hidden dependencies (LeCun et al, 2015).

On well-established machine learning approach is he Recurrent Neural Network (RNN), which is particularly effective for tasks involving temporal dependencies such as time forecasting. However traditional RNNs suffer from limitations, including vanishing gradients and difficulty in capturing long-range dependencies (Bengio et al, 1994).

A more recent and highly efficient deep learning approach is the transformer model (Vaswani et al, 2017), which replaces recurrence with a self-attention mechanism. This allows the model to process entire sequences simultaneously rather than sequentially, making it more scalable and effective for capturing long-range dependencies. These transformers have achieved state-of-the-at results in various fields, including natural language processing and time-series prediction (Wen et al, 2022).

In the agricultural domain, RNNs have been applied to classify crops using satellite imagery (Ndikumana et al, 2018), and transformers have been developed to enhance crop yield predictions (Bi et al, 2023). However, the use of deep learning techniques, particularly the use of transformers-based models, to predict crop rotations based on historical sequences remains largely unexplored. Applying such a model to predict farmers' decision-making and simulate crop rotation patterns would thus address a technological gap.

1.3 Aim and research questions.

This thesis aims to explore the potential of predicting farmers' decision-making using a dataset detailing previous crop cultivation per parcel. The first objective is to analyse existing crop rotation patterns and their response to cover crop policies through a data-driven approach. Additionally, it explores the application of emerging machine learning techniques to simulate aspects of the farmers' decision-making. Creating a foundation for a model that policymakers could use to simulate the effect of policy changes.

To reach these objectives, this thesis follows two complementary approaches. At first a data science-driven approach, analysing major crop rotation clusters using an edit distance matrix and secondly a machine learning approach, where a transformer model is trained to predict crop rotations auto-regressively. The model's outputs will be compared with baseline distribution and validated against the established crop rotation clusters to assess its predictive capabilities.

The research questions guiding this investigation are as follows:

RQ 1: What is the influence of crop rotations on the presence of cover crops in the Netherlands?

> **SRQ 1.1:** To what extent do the detected clusters represent common crop rotations in the Netherlands?
>
> **SRQ 1.2:** How do these clusters relate to cover crop presence between 2017 and 2023?
>
> **SRQ 1.3:** How do different regions in the Netherlands compare in terms of main crop rotations and cover crop use?

RQ 2: To what extent can a transformer-encoder model be used to predict future crops, while solely being trained on previous crop rotations?

Through these questions, this thesis aims to bridge the gap between data science and agriculture. This is done by exploring and combining known data analysis methods from other fields, as well as providing an innovative view of how machine learning can be leveraged to predict future crop rotations and ultimately assist policymaking.

2 Data and methods

The research questions will be addressed through two primary approaches: a data science method and a machine learning method. The data science approach explores the dataset by applying hierarchical clustering to a subsample dataset of crop rotations in the Netherlands. The machine learning approach utilizes a transformer model to predict future crops based on the crops planted previously on a field. A general overview of both approaches is shown in the figure below (Figure 1).

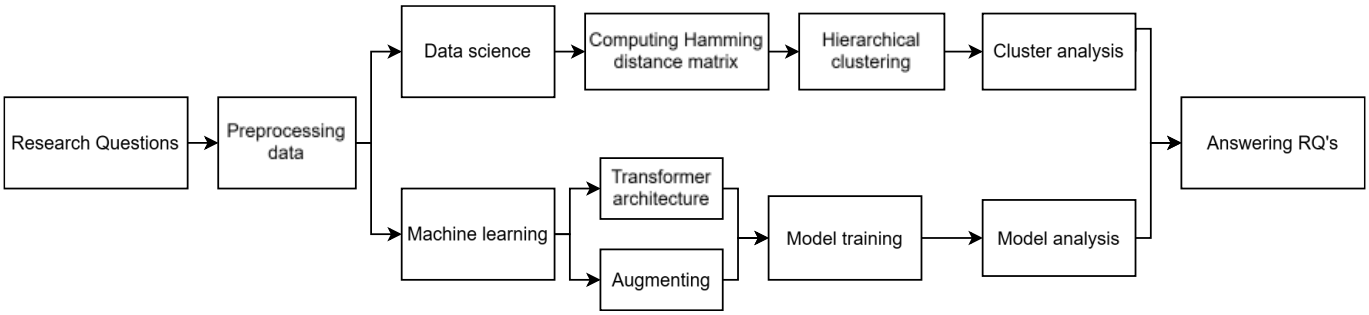


Figure 1: Flowchart of the methodology

2.1 Data & study area

For this thesis, an extensive field parcel dataset from the Dutch government, the "*Basisregistratie Gewaspercelen (BRP)"* has been used. This dataset is constructed from farmers' declarations, which are collected annually through self-reports. The BRP was requested and made available for research purposes. It comprehends data on field location, labelled field boundaries, soil types and planted crops per year. Initially, the dataset was structured as separate yearly records. To compile a single crop rotation dataset, field polygons were matched across years using 2023 as the reference. For each field polygon in the 2023 dataset, the corresponding fields from previous years were identified, and the crop label was assigned based on the largest intersecting polygon. This compiled dataset includes all agricultural fields in the Netherlands from 2017 to 2023. As this research focuses on rotation practices between different crops, all fields which are marked as permanent or natural grass were excluded. To prevent problems introduced by missing farmer declarations, all parcels in which the main crop column was left empty for one or more years, were excluded from the dataset. This kept consistency in crop rotation lengths over the dataset. As the dataset was also pre-processed for use with satellite images, all fields with an area smaller than 0.5 hectares were removed, resulting in approximately 105,000 fields remaining in the dataset.

2.2 Preprocessing sequence

The initial step in preparing the dataset involved grouping certain crops based on their functional similarity; for example, different types of sugar beet within the BRP were consolidated into a single category. The crop groups used in this research are shown in Figure 2, which also shows their frequency in the BRP dataset. A more specific set of crops, 26 in total, has been chosen compared to the ten crop groups considered by Ballot et al. (2023), as one of the aims is to analyse and predict the exact crop rather than a broad crop group. The resulting sequences all have a length of seven crops, with each position representing a specific year between 2017 and 2023.

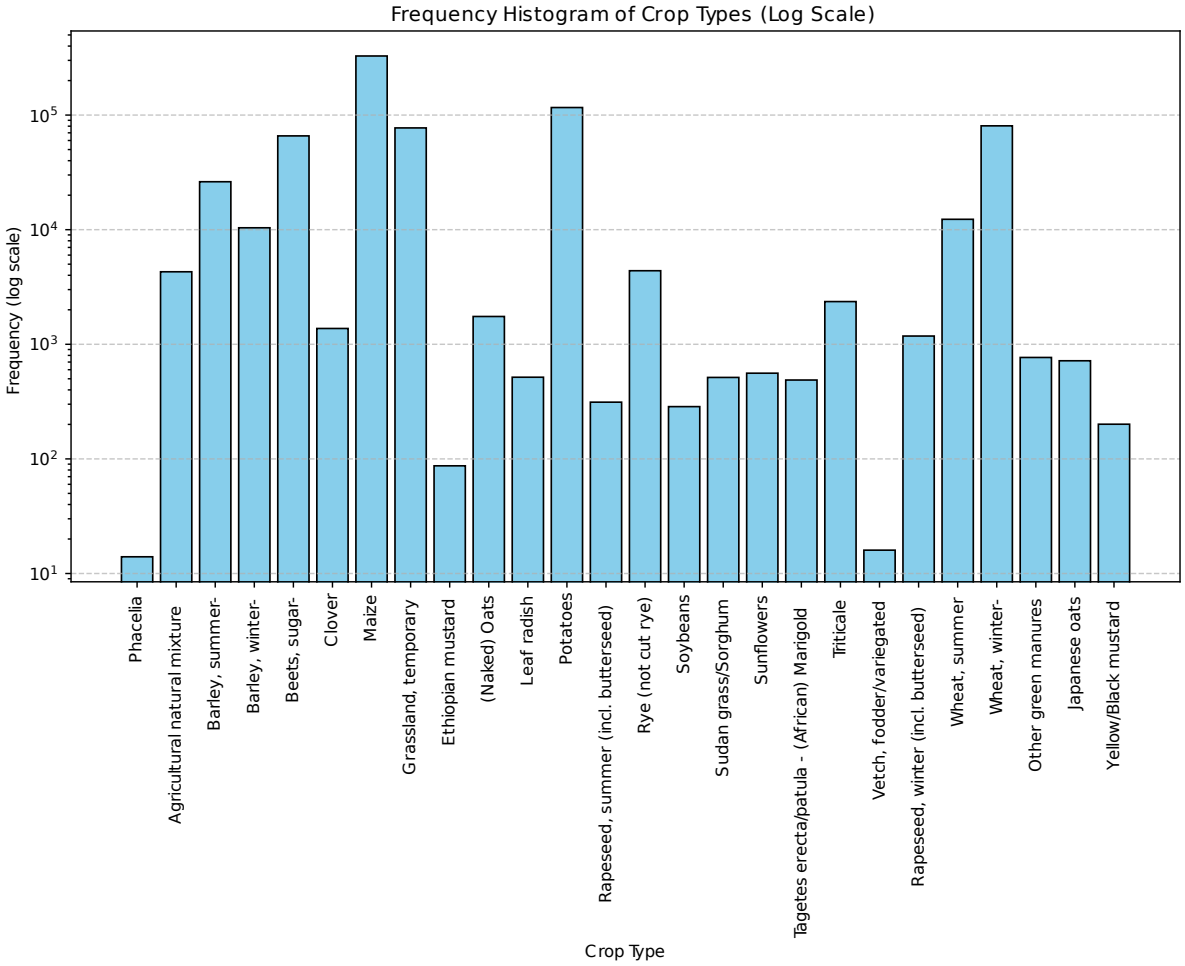


Figure 2: The frequency of occurrence per crop category used in this research

2.3 Data science approach

2.3.1 Preprocessing

Due to computational reasons, a subsample of the full dataset was utilized for the data science approach. To ensure the representativeness of the subsample, two conditions were defined: First, each province needed to be represented by a sufficient number of fields to enable regional comparisons. To achieve this, the threshold was set to ensure an equal distribution of fields across provinces, with each province contributing to 1/12 of the dataset. Secondly, the distribution of the main clusters had to remain stable. Subsamples of 1200, 3000, 6000 and 12000 fields were tested, revealing that the cluster distribution stabilized between 6000 and 12000 fields. Consequently, a subsample of 12000 fields was selected for the data science approach.

2.3.2 Hamming distance matrix

To properly cluster crop rotation sequences, a metric to calculate the similarity between individual sequences is needed. Since crop types are categorical, and thus should not be interpreted numerically, the Hamming distance (Hamming, 1950) was chosen. This distance quantifies the edit distance between two sequences. A matrix is obtained by applying it to each sequence, measuring the number of positions in the sequence which differ compared to all other sequences. Mathematically the hamming distance is calculated as follows:

$$d_H(X,Y) = \sum_{i=1}^{n} 1(x_i \neq y_i)$$

With $X = (x_1, x_2, \dots, x_n)$ and $Y = (y_1, y_2, \dots, y_n)$ representing the sequences that are compared and n being the sequence length. According to this formula, the hamming distance is the sum of every time $x_i \neq y_i$. A lower Hamming distance indicates a greater similarity between the compared sequences. For example:

$$d_H(Maize > Rye > Wheat, \quad Maize > Grass > Sugar\ beet) = 2$$

Indicates a higher dissimilarity then:

$$d_H(Maize > Rye > Wheat, \quad Maize > Rye > Wheat) = 0$$

However, this traditional Hamming distance approach doesn't account for the fact that crop rotations do not necessarily have to start in the same year. For example:

$$d_H(Wheat > Grass > Maize > Rye, \quad Grass > Maize > Rye > Wheat) = 4$$

These sequences basically follow the same crop rotation, however when applying the hamming distance formula to these sequences it returns the value 4, which incorrectly suggests complete dissimilarity. To prevent this from happening we introduce the cyclic hamming distance:

$$d_H^{cyc}(X,Y) = \min_{s \in \{0,\dots,n-1\}} d_H(X, shift(Y,s))$$

Where shift(Y,s) cyclically shifts the sequence Y by all s positions, without altering the order, before calculating the Hamming distance. The resulting minimal hamming distance is saved in the matrix. In the previous case this would result in:

$$d_H^{cyc}(Wheat > Grass > Maize > Rye, \quad Grass > Maize > Rye > Wheat) =$$

$$d_H(Wheat > Grass > Maize > Rye, \quad shift(Grass > Maize > Rye > Wheat, 0) = 4$$

$$d_H(Wheat > Grass > Maize > Rye, \quad shift(Wheat > Grass > Maize > Rye, 1) = 0$$

$$d_H(Wheat > Grass > Maize > Rye, \quad shift(Rye > Wheat > Grass > Maize, 2) = 4$$

$$d_H(Wheat > Grass > Maize > Rye, \quad shift(Maize > Rye > Wheat > Grass, 3) = 4$$

Where 0 is the lowest Hamming distance found, correctly suggesting that both sequences are the same crop rotation and should thus be classified as similar.

2.3.3 Hierarchical clustering

We now have a matrix which quantifies the diversity in crop rotations, to analyse and group these patterns agglomerative hierarchical clustering (Jain & Dubes, 1988) will be introduced below. This recursive merging process works as follows:

1. First, each data point in the Hamming distance matrix is initialized as its cluster.

2. Then a pairwise distance matrix D is computed, where the distance. D_{ij} Represent the dissimilarity between clusters. C_i and C_j.

   Mathematically this distance between two clusters C_i and C_j Is computed as:

$$D(C_i, C_j) = \frac{1}{|C_i||C_j|} \sum_{x \in C_i} \sum_{y \in C_j} d_H^{cyc}(X, Y)$$

Where $|C_i||C_j|$ represents the size of clusters C_i and C_j, by dividing by $|C_i||C_j|$ The formula ensures that the distance measure remains independent of cluster size, making it possible to compare across different pairs of clusters.

3. Then the closest clusters are merged, in our case it will merge every sequence that is the same and thus have a hamming distance of 0 between all sequences. This merging process is encoded as follows in a linkage matrix:

$$Z = [i \quad j \quad d_{ij} \quad n_{ij}]$$

Were i and j are the merged clusters, d_{ij} Is the distance between them and n_{ij} Is the number of original points in the new cluster.

4. This process is repeated using the newly merged clusters as input, and continues until the set threshold of 3 is met:

$$D(C_i, C_j) > 3$$

This prevents clusters from being formed with a linkage distance greater than 3, a number chosen to balance similarity within clusters while avoiding excessive fragmentation of clusters and thus crop rotations.

To visualize the concept of hierarchical clustering and the influence of the chosen threshold, a dendrogram was constructed. Figure 3 shows an example of 30 random crop sequences, clustered using a Hamming distance matrix. The dendrogram shows that setting the threshold at 0, returns 18 clusters out of 30 sequences, as some sequences are identical to each other. While the threshold at 3 will result in 8 clusters.

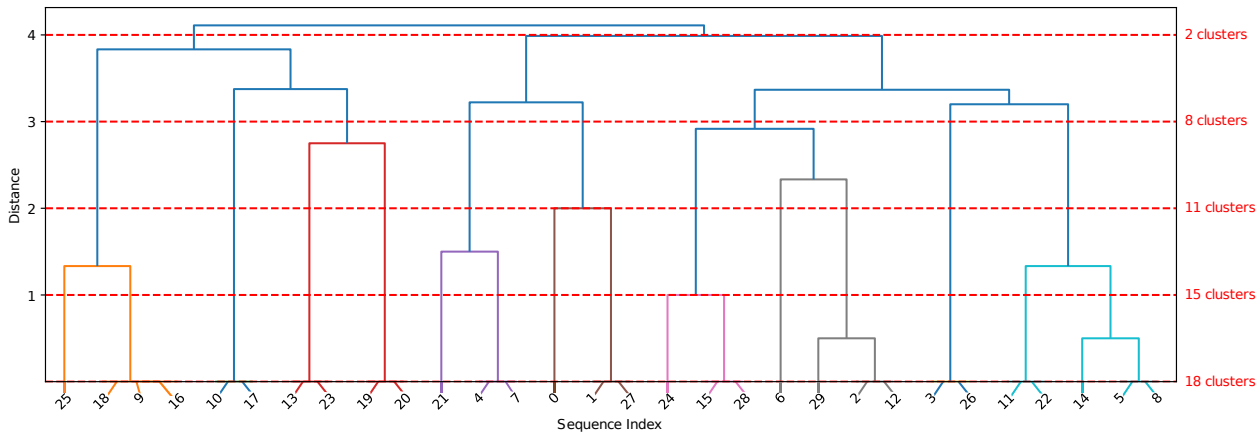


Figure 3: An example of clustering using a dendrogram, showing the importance of setting the right threshold

2.3.4 Cluster analysis

Each sequence was assigned to a specific cluster, enabling a wide range of analyses on relationships between common crop rotations, soil types, regions, crop diversity, and cover crop usage. Each cluster was visually inspected, described, and given a distinct name. To spatially compare these variables, NUTS 3 regions, were employed. These regions are part of the Nomenclature of Territorial Units for Statistics (NUTS), a system used by the EU for regional classification, where the NUTS 3 represent small regions within a country. Next the comparing these regions a visual comparison of the results and the 12 provinces of the Netherlands was executed. Crop diversity has been measured by categorizing fields into three levels: *simple*, *moderate* and *diverse*. Fields have been classified as *simple* if, between 2017 and 2023, less than three unique main crops have been cultivated. *Moderate* fields had exactly three unique crops, and *diverse* fields saw more than three unique crops over these seven years. This diversity metric serves as an indicator of the prevalence and extent of crop rotation practices across different provinces.

2.4 Machine learning approach

2.4.1 Preprocessing

Since most farmers cultivate multiple fields, there is a risk of data leakage between the training and validation datasets. This could occur if farmers apply similar crop rotation practices across different fields, leading to a potential overlap of practices between both datasets. To minimize spatial autocorrelation, a gridded data split has been executed. Firstly, the Netherlands was divided into 5x5 km grid cells. Under the assumption that most farmers cultivate fields that are in close proximity to each other, this assigns most of the fields that are owned by a single farmer to the same grid cell. These grid cells were then used to split the dataset into training, validation and test sets following a 70/15/15 ratio.

2.4.2 Transformer model

A Transformer Encoder (Vaswani et al., 2017) is implemented and adjusted, a schematic overview of this model is illustrated in Figure 4. Paragraph *2.4.3* will further expand on the key components of the model's architecture.

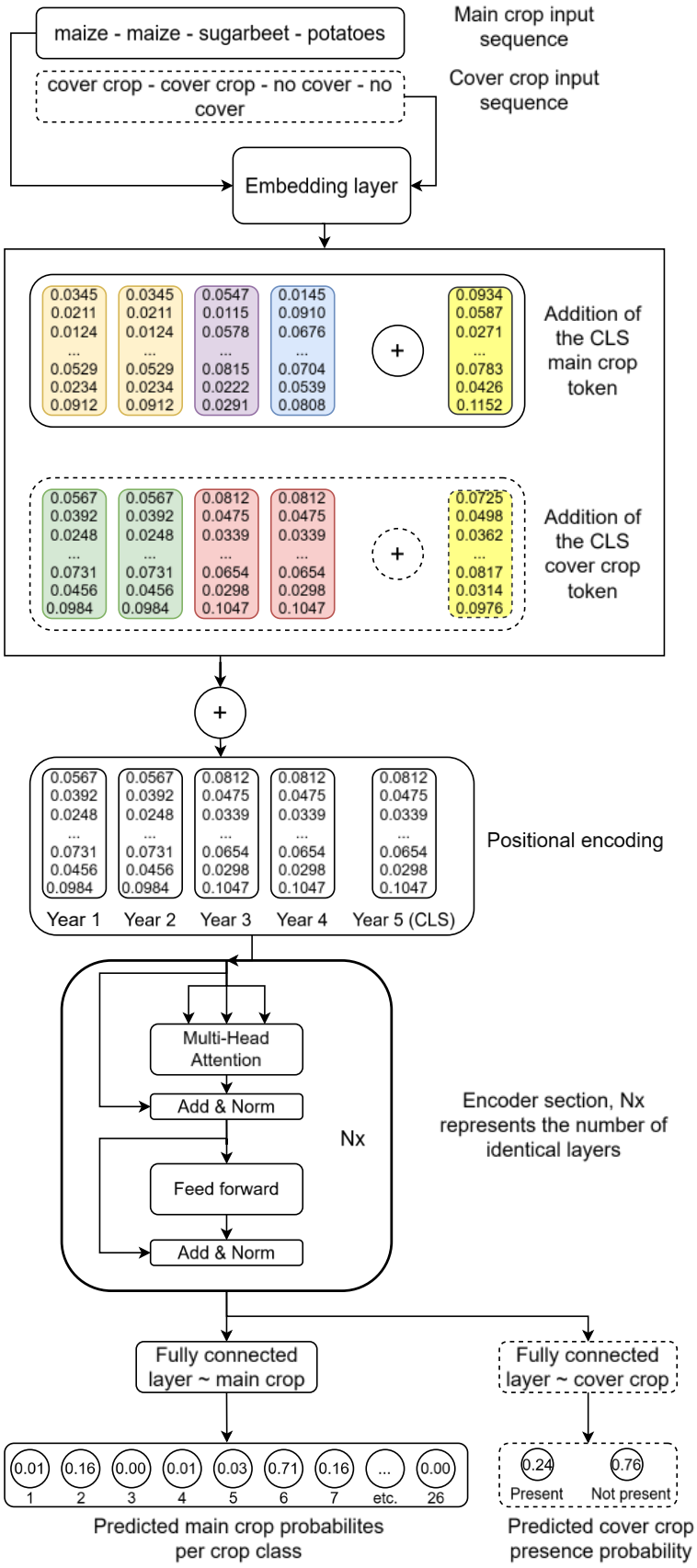


Figure 4: The transformer-encoder model architecture.

2.4.3 Model architecture

2.4.3.1 Embedding layer

The embedding layer is the initial layer of the transformer model architecture. This layer converts discrete input tokens, crop-type classes in this case, into continuous vector representations. Each crop type is assigned a unique vector representation, which is updated at each forward pass, as the model captures the semantic relationships between different crop types.

2.4.3.2 Learnable Classification (CLS) tokens

The second layer introduced in the transformer model regulates the addition of learnable CLS tokens, which are added at the end of each sequence. In the crop rotation model, two distinct CLS tokens are utilized: one for predicting the main crop and another for predicting the presence of cover crops. Although predicting cover crop presence is beyond the scope of this thesis, its implementation facilitates future research on cover crop modelling.

CLS tokens are appended at the end of each input sequence and initialized as learnable parameters. During training, the model updates these tokens to aggregate contextual information from the entire sequence through its self-attention mechanisms (*2.4.3.4*)

After passing the CLS token through all layers of the transformer, the final hidden state is used as output to predict the following crop. By passing the CLS token through a fully connected layer, it is transformed into a vector of prediction probabilities for each crop class.

2.4.3.3 Positional encoding

Since a traditional transformer model does not inherently consider the order of sequences during training, positional encoding has been introduced to capture the temporal dependencies of crop rotation sequences. This has been achieved by using sinusoidal positional encoding, which operates as follows:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{1000^{\frac{2i}{d_{model}}}}\right)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{1000^{\frac{2i}{d_{model}}}}\right)$$

With $PE_{(pos,2i)}$ representing the positional encoding for even indexes, each index being represented by pos, and $PE_{(pos,2i+1)}$ representing the positional encoding for uneven indexes. d_{model} is the total embedding dimension and $1000^{\frac{2i}{d_{model}}}$ Is the scaling factor.

2.4.3.4 Attention mechanism

One of the key components of the transformer architecture is the multi-head-self-attention mechanism. It enables the model to learn temporal patterns and contextual information between consecutive crops within the crop rotation sequences. Self-attention computes per element within the sequence the importance relative to all other elements. With $X = (x_1, x_2, \dots, x_n)$ representing the input sequence, as introduced by Vaswani, the attention scores are computed as follows:

First, each element is projected into three vectors: Query (Q), Key (K) and Value (V), using learned weight matrices:

$$Q = XW^Q, \qquad K = XW^K, \qquad V = XW^V$$

Where W^Q, W^K and W^V represent the weight matrices for the query, key and value. The matrix of outputs is then computed using:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_K}}\right)V$$

Where d_K is the dimensionality of K, introduced as a scaling factor to maintain stable gradients. To enhance the model's ability to learn different aspects of the given crop rotation sequence, the multi-head-self-attention mechanism is introduced. This extends the single attention layer by employing multiple attention heads in parallel.

$$MultiHead(Q, K, V) = Concat(head_1, head_2, \dots, head_h)W^O$$

Where each head i is computed through:

$$head_i = Attention(QW_i^Q, QW_i^K, QW_i^V)$$

These functions h represent the number of attention heads, and W_i^Q, W_i^K, W_i^V *and* W^O represent the weights matrices for each head.

2.4.4 Model training

2.4.4.1 Augmentation

The following preprocessing step introduces variable sequence lengths as a form of random data augmentation. Figure 5 explains how augmentation is applied to each sequence. First, three random lengths between 3 and 7 are chosen per individual sequence. Then, the original sequence is truncated for those 3 lengths, starting at a random index. Next to tripling the size of the dataset, the process of incorporating variable sequence lengths improves the ability of the model to learn more about the complexity and diversity of crop rotation patterns. Augmenting the data simulates the dynamic nature of farming practices, where fields may be managed by different farmers introducing different rotation practices over time.

2.4.4.2 Class weighting & loss function

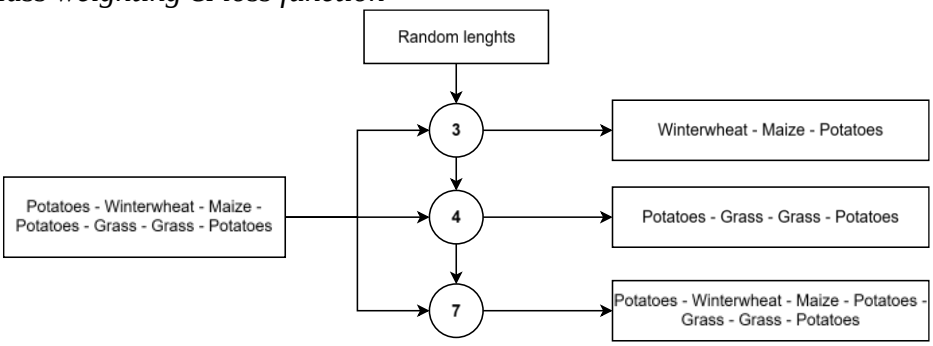


Figure 5: Augmenting the dataset by introducing variable sequence length

As the dataset is imbalanced, due to common crops like maize and sugar beets occurring far more frequently than some less common crops, balanced weighting is applied to prevent the model from overpredicting on major crop classes. The weighting will skew the bias more towards minority classes. These weights are computed as follows:

$$w_i = \frac{n}{n_i \cdot C}$$

With w_i representing the weight of class i. n representing the total number of crops planted over the dataset, and n_i being the total number of crops of class i planted. C represents the number of crop classes in the dataset.

These weights are applied within the loss function, which in the case of this multi-class classification problem is the Cross-Entropy-Loss function, which combines the SoftMax activation and negative log-likelihood loss:

$$L = -\sum_{i=1}^{C} w_i y_i \, log(\hat{y}_i)$$

Where L is the total loss of a single sample, C is the number of classes, w_i is the weight of the class i. y_i is the true label for class I and is one-hot encoded, so $y_i = 1$ for the correct class and 0 otherwise. $\hat{y}_i$ is the predicted probability of class i according to the output of the SoftMax function.

2.5 Model analysis

2.5.1 Model evaluation

The model is trained on a trial-and-error approach, where hyperparameters affecting model complexity, training duration and learning dynamics were systematically adjusted. The training objective was to minimize validation losses while achieving relatively high evaluation performances. In addition to these training metrics, the model was tested & evaluated using three approaches: a *quantitative* comparison of output distribution across three sequence groups and variable length, a *qualitative* assessment of the output for individual model predictions, and an assessment of the model's ability to *reconstruct* clusters from scratch.

2.5.1.1 Computing equal distribution (ED), crop distribution (CD) & reference distribution (RD)

To derive meaningful insight from the output distribution(s), four methods were used to construct comparative baseline distributions. The equal distribution (ED) represents the most basic distribution and assumes no prior knowledge of agricultural practices by using a uniform distribution. The crop distribution (CD) is based on the frequency of each crop class within the dataset. This baseline incorporates knowledge of common crop occurrences in the Netherlands, however it does not account for crop rotation patterns. The references (RD) correspond to the distribution of the true labels, reflecting the actual distribution which the model is aimed to predict. As an example, the RD, ED & CD of the next crop in the sequence sugar beets – winter wheat – potatoes – winter wheat is illustrated in Figure 6

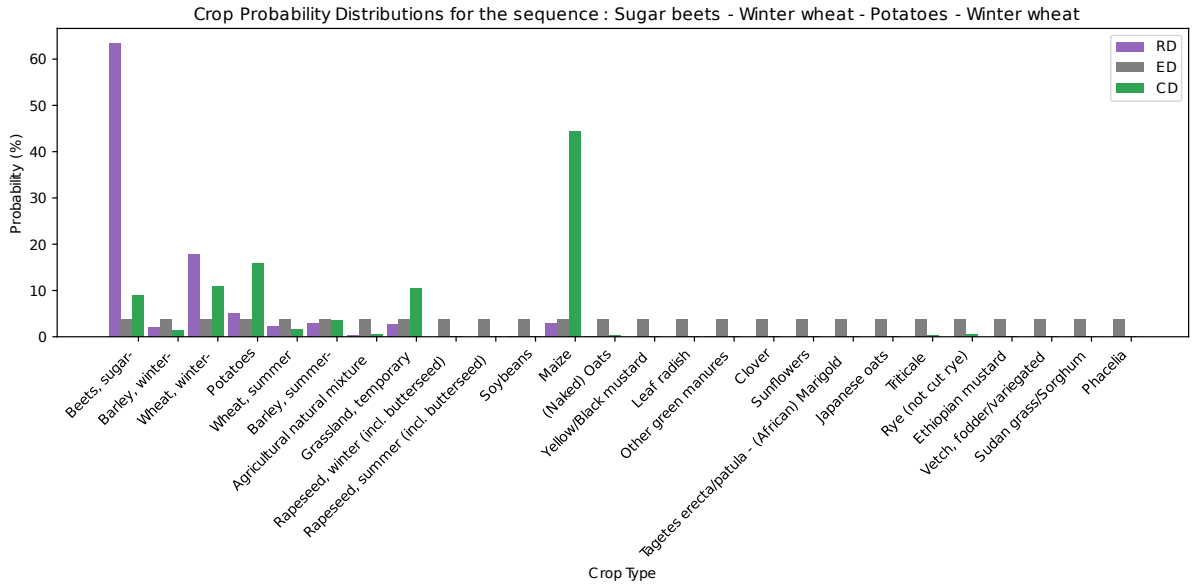


Figure 6: An example of RD, ED & CD.

2.5.1.2 Computing the lookup distribution (LD)

The lookup-based distribution (LD), is a data-based approach to approximate crop prediction. It operates as an extensive lookup table, scanning the entire 2017-2023 field database, abbreviated with α, to look for sequences similar to the input. The computation of LD begins with a sequence of at least two years. Initially, the first crop in the input sequences S_i is removed, effectively shifting the sequences one year back in time. This is needed as the concept of LD is based upon the idea that less information is available, as it looks back in time:

$$S_i' = (S_i[2]S_i[3], \dots, S_i[\mathrm{n}])$$

Thus, S_i' represent the sequences obtained after removing the first crop $S_i[1]$. In practice, this will look like this:

$$S_i = Wheat, Grass, Maize, Rye$$

$$S_i' = Grass, Maize, Rye$$

The next step involves searching within the full sequence dataset, α, for sequences that exactly match the first n crops S_i'. This set of matching sequences $\delta_{S_i'}$, is computed as follows:

$$\delta_{S_i'} = \{S_j | S_j \in \alpha, S_j[1{:}n] = S_i'[1{:}n]\}$$

Here, n represents the length of S_i', meaning that $S_j[1{:}n]$ refers to the first n crops of any sequence in dataset α. For example, searching S_i' in dataset α could return:

$$\delta_{S_i'} = (Grass, Maize, Rye, Maize), (Grass, Maize, Rye, Maize),$$
$$(Grass, Maize, Rye, Barley), (Grass, Maize, Rye, Rye), (Grass, Maize, Rye, Rye),$$
$$(Grass, Maize, Rye, Grass), (Grass, Maize, Rye, Grass), (Grass, Maize, Rye, Grass)$$

The selected set of sequences, $\delta_{S_i'}$, is then used to investigate the statistical likelihood of farmers continuing a crop rotation that exactly matches S_i'.

In this study, 26 crop classes are considered, each assigned a numerical identifier ranging from 0 to 25, denoted by k. The term $crop_{n+1}^{k}$ represents the occurrence k-the crop, where $k \in \{0,1, \dots,25\}$, in the $(n+1)$-th position across all sequences $\delta_{S_i'}$. The LD distribution can then be computed as follows:

$$LD(crop_{n+1} = k \mid \delta_{S_i'}) = \frac{count_k}{|\delta_{S_i'}|} \quad for\ each\ k \in \{0,1, \dots,25\}$$

Where $count_k$ represents the number of sequences in $\delta_{S_i'}$ where the $(n+1)$-th crop is k.

In our example LD would be the distribution along the 4th position in the selected sequences in $\delta_{S_i'}$:

$$LD(crop_{n+1} = k \mid \delta_{S_i'}) = (0.375, 0.250, 0.250, 0.125, 0.000 \dots 0.000)$$

Thus, in the example, LD predicts a 37.5% chance that the next crop in S_i will be grass, followed by maize and rye at 25% each, and barley at 12.5% The remaining 22 crops will not be planted (0%) according to LD.

2.5.2 Comparing the influence of length and frequency on the predicted distribution

First, the model's output will be compared over different sequence length and frequency groups, to spot trends over the entire dataset. For this broader evaluation of the model, the Kullback-Leibler (KL) divergence is used as a comparison metric. This measure is introduced by Kullback and Leibler (1951) and quantifies the difference between the predicted probability distribution of the next crop in the rotation and the actual probability distribution. The KL-divergence indicates how the predicted distribution *Q* deviates from the true distribution *P* and is computed as follows:

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

Where $P(i)$ is the true distribution, $Q(i)$ is the approximated distribution. The more the $D_{KL}(P||Q)$ value deviates from 0 the bigger the model's predicted distribution deviates from the ground truth. In Figure 7, two KL divergences are visualized. The left distribution exhibits similar probability values for both P and Q, resulting in a relatively low KL divergence (0.0002). In contrast, the right distribution shows greater dissimilarity between P and Q, leading to a higher KL divergence of 0.5558.

To contextualize the divergence values between the model's output distribution (TD) and the actual distribution (RD), the divergences between the RD and the three baseline distributions, as introduced in section *2.5.1*, are also computed. To generate an insight into the effect of how often a crop rotation is applied, three frequency groups are introduced based on the test set: *common* (over 250 occurrences), *uncommon* (between 50 and 249 occurrences), and *rare* (between 10 and 49 occurrences). The input will be of variable sequence lengths, creating the possibility of investigating the effect of different lengths.

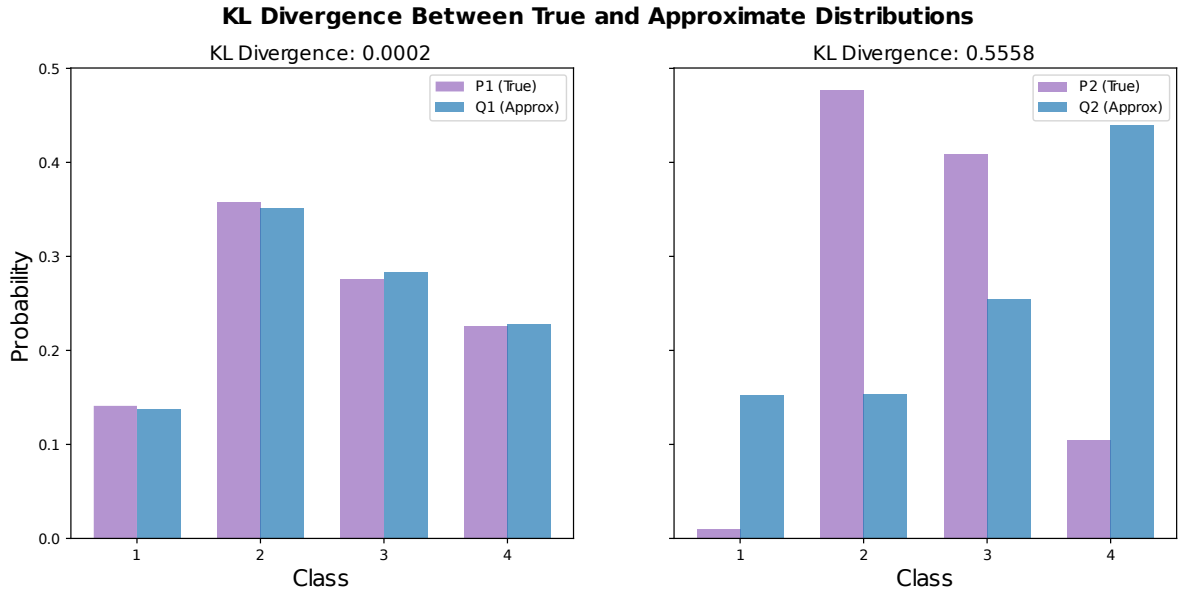


Figure 7: An example comparison of two probability distributions and the computed KL-divergence, showing that a similar distribution results in a lower KL-divergence.

2.5.3 Investigating the prediction performance of individual sequences

For the qualitative analysis, two sequences were selected for comparison. These sequences were chosen based on the results of the broader comparison introduced in the previous paragraph and should give insight into how TD as well as ED/CD/LD handle the comparison of a single sequence. The output distribution of the model (TD) for these sequences was compared against 1) the actual distribution (RD) and 2) the weighted crop class distribution (CD). Additionally, the KL-divergence, as explained in *2.5.2*, was computed between *RD & TD* as well as between *RD & CD*, to quantify the differences between these distributions.

2.5.4 Reconstructing sequences

The final model evaluation measure used is sequence reconstruction. In this approach, the model is provided with a single initial crop from a random sequence in the dataset. The model then predicts the top three most probable second crops, which are subsequently used to generate three possible third crops. This process is repeated iteratively until a sequence length of 7 is reached, resulting in a total of $3^6 = 729$ possible reconstructed sequences per input crop. The evaluation assesses whether the original sequences are eventually present within those 729 sequences. This procedure has been conducted for a total of 1000 randomly selected sequences.

For example, we examine whether the common sequence of seven years of continuous maize cultivation can be reconstructed. The model first receives maize as the initial input and then predicts the next crop, returning the top three probabilities:

$$Top3\ (maize) > maize, sugar\ beet, grass$$

Then, these predictions are iteratively used as input, leading to:

$$Top3\ (maize - maize) > maize, grass, potatoes$$

$$Top3\ (maize - sugar\ beet) > winter\ wheat, potatoes, maize$$

$$Top3\ (maize - grass) > grass, maize, potatoes$$

This process continues until a sequence of length 7 is constructed. Subsequently, all generated sequences are evaluated to determine whether the original sequence, in this case, a maize monoculture, has been accurately reconstructed.

3 Results

3.1 Clustering

This section presents the results of the hierarchical clustering of the sample dataset. Next to dividing this dataset into clusters, it aims to examine the relationship between crop rotation, cover crop choices and regional influences in the Netherlands. First, the identified major crop rotation clusters will be compared to common farming practices. Then, the clusters will be compared in terms of cover crop usage between 2017 and 2023. Finally, spatial relationships will be outlined and linked to clusters, soil texture and cover crop usage.

3.1.1 Detecting & describing main crop rotation clusters

Over 400 clusters were identified in the subsample dataset using hierarchical clustering. To focus on the most common crop rotations, a minimum cluster size threshold of 50 fields was applied, corresponding to approximately 830 fields across the full dataset. The resulting distribution of these clusters is shown in Figure 8. Clusters that do not meet the set threshold are categorized as “other” and are not considered a major crop rotation in this thesis. A detailed list of all 18 major clusters found, including their assigned cluster numbers, names and descriptions, can be found in Table 1.

A notable finding is the variation in cluster sizes among the major crop rotations. The largest identified cluster, cluster 44, is described as a maize monoculture cluster. This cluster mainly includes sequences where maize has been cultivated almost continuously for the past seven years, which is a common practice in dairy-dominated areas in the Netherlands (Velthof et al., 2020). This regional influence will be further explored in paragraph *3.1.3*. Next to maize, grass is a common fodder crop for dairy farmers (Schils et al., 2002), clusters 56, 185 & 186 reflect this practice as these rotations alternate dominant grass with common crops like maize and potatoes.

Another common crop rotation strategy involves alternating potato cultivation every three to four years to mitigate the risk of plant diseases (Johnson & Dung, 2010). This rotation is exemplified in clusters 265 & 266, where potatoes, sugar beets and winter wheat follow a mostly structured rotation. Clusters 324 & 336 employ a similar rotation strategy, replacing winter wheat with maize.

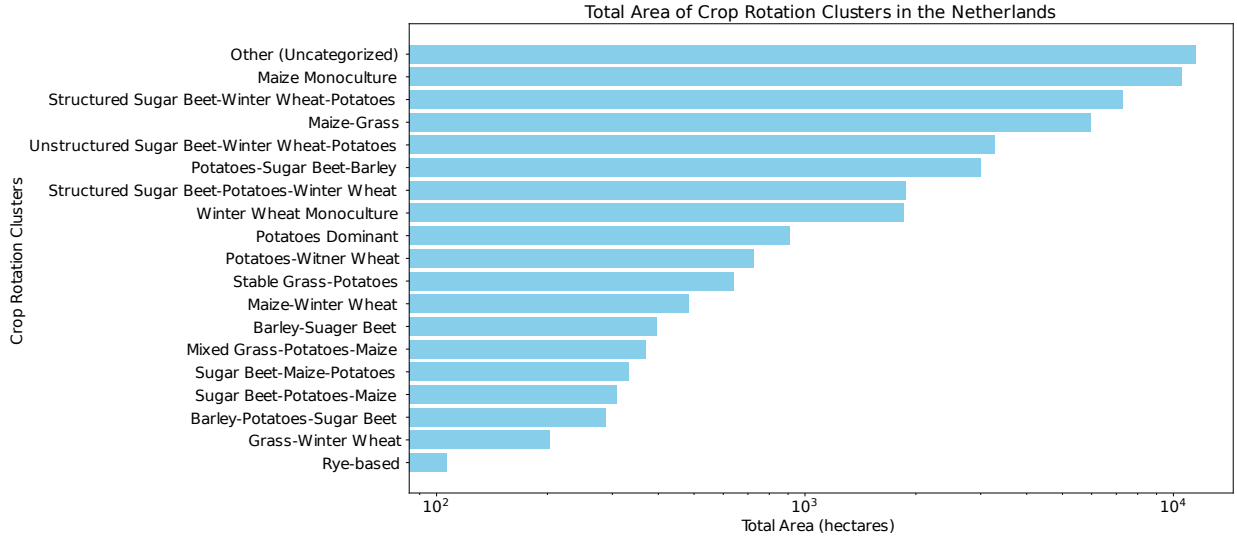


Figure 8: Major crop rotation cluster distribution

Table 1: Major cluster numbers, names and descriptions

Cluster	Name	Description
8	Rye-based rotation	Cluster with dominantly rye sown. Sometimes alternated with other cereals, like winter wheat, triticale or summer barley.
44	Maize monoculture	A cluster consisting of a full maize monoculture, occasionally a deviation of 1 crop over 7 years.
56	Maize-grass rotation	A cluster of fields with a maize-grass rotation, where maize and grass both have been cultivated at least twice over 7 years.
185	Mixed grass-potatoes-maize rotation	A cluster where predominantly grass and potato are cultivated, but rotated inconsistently and often alternated with different crops.
186	Stable grass-potatoes rotation	Cluster with a consistent rotation between grass and once in the 3-4 years potatoes. At times grass is swapped out for maize.
197	Grass-winter wheat rotation	Cluster where grass and winter wheat rotate over the 7 years.
238	Maize-winter wheat rotation	Cluster where maize and winter wheat alternate each other inconsistently. Occasionally another crop pops up once.
265	Unstructured sugar beet-winter wheat–potatoes rotation	Cluster with a diverse rotation pattern of sugar beet, winter wheat and potatoes. Where all three crops are present in similar proportions.
266	Structured sugar beet-winter wheat-potatoes rotation	Cluster with a structured rotation pattern of sugar beet, winter wheat followed by potatoes.
267	Potatoes-winter wheat rotation	Cluster where winter wheat and potatoes predominantly rotate.
269	Structured sugar beet-potatoes-winter wheat rotation	Cluster with a structured rotation pattern of sugar beet, potatoes followed by winter wheat.
275	Winter wheat monoculture	Cluster with a dominant cultivation of winter wheat. On all fields, winter wheat occurs at least 5 times.
324	Sugar beet-potatoes-maize rotation	Cluster with a structured rotation pattern of sugar beet, potatoes followed by maize. Occasionally completed with a cereal like winter barley or winter wheat.
336	Sugar beet-maize-potatoes rotation	Cluster with a structured rotation pattern of sugar beets, maize followed by potatoes.
355	Potatoes dominant rotation	Cluster where potatoes at cultivated at least 3-4 times, commonly rotated with winter wheat and/or sugar beets, with occasional inclusion of another crop.
361	Barley-potatoes-sugar beet rotation	Cluster with a structured rotation pattern of barley, and potatoes followed by sugar beets. Interestingly, this rotation often includes two consecutive years of potatoes before continuing with sugar beets.
364	Barley-sugar beet rotation	Cluster with barley as the dominant crop, mostly followed by sugar beets, with potatoes incorporated into the rotation once every 7 years
371	Potatoes-sugar beet-barley rotation	Cluster with a high frequency of potatoes rotated with sugar beets and summer barley.

3.1.2 The influence of clusters on the use of cover crops

Figure 9 illustrates the distribution of cover crop frequency, and shows how the majority of the fields within the dataset have practised the use of cover crops in the 2017-2023 period. The frequency of cover crop usage differs strongly when comparing major clusters (Figure 10).

For example, in the dominant cluster 44, representing maize monoculture, the frequency of cover crops is relatively high. A similar effect can be observed in cluster 238, which is also dominated by maize cultivation.

In contrast to the "maize" clusters, grass-dominated clusters (56, 185 & 186) are characterized by a lower frequency of cover crop usage. This can be explained by the fact that grass is cultivated throughout winter time, reducing the necessity for cover crops. Another interesting exception in the overall distribution is cluster 8, representing fields of rye monoculture. Since rye is a winter crop, and thus sown in autumn, there is no need for a cover crop during the winter period.

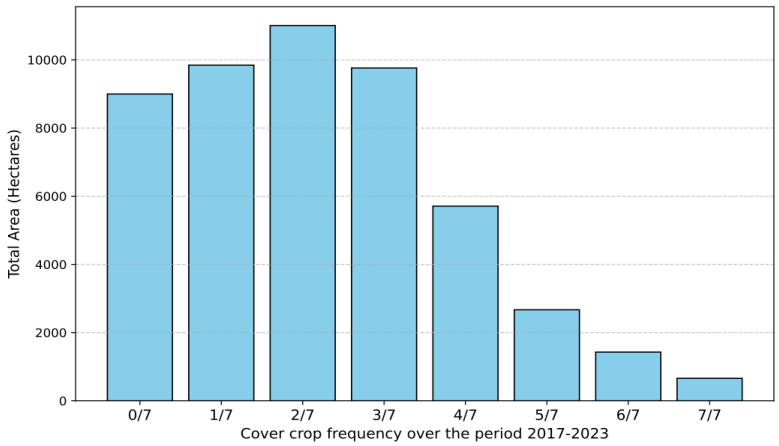


Figure 9: Distribution of cover crop frequency over the total area in the sample dataset. The fractions represent the number of cover crop declared over the 7-year period.

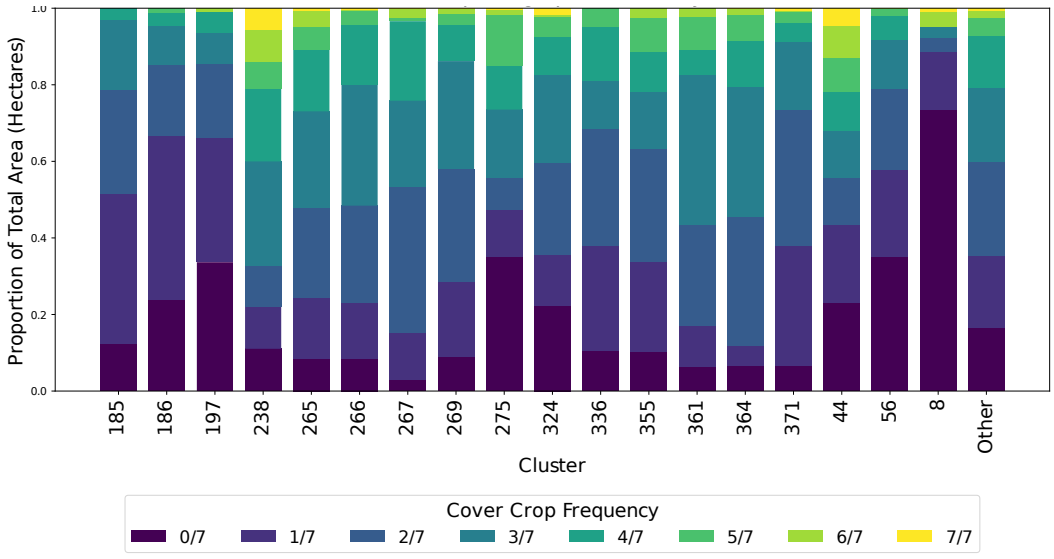


Figure 10: Cover crop usage frequency per major crop rotation cluster

3.1.3 Mapping spatial relations between major clusters and cover crops

Investigating the spatial relationship between clusters, soil composition emerges as a key factor influencing agricultural practices. Figure 13c shows the soil texture across the Netherlands, illustrating that the northern and western provinces are predominantly characterized by loamy soils, whereas the eastern and southern regions are primarily composed of sandy-loamy soils. With the exception of southern Limburg, which is classified as a silt-loam area.

Figure 11 compares the provinces by crop diversity, showing there is quite a difference in cropping diversity between the provinces. One such correlation between soil texture and crop diversity can be observed. Provinces with predominantly loamy soils, such as Flevoland and Zeeland, exhibit a higher-than-average crop diversity. In contrast, Overijssel, Noord-Brabant and Gelderland, the sandy-loam soils, demonstrate a less diverse cropping pattern. An interesting exception is Utrecht, a predominantly clayey province, which records the lowest crop diversity. Figure 13b highlights this spatially.

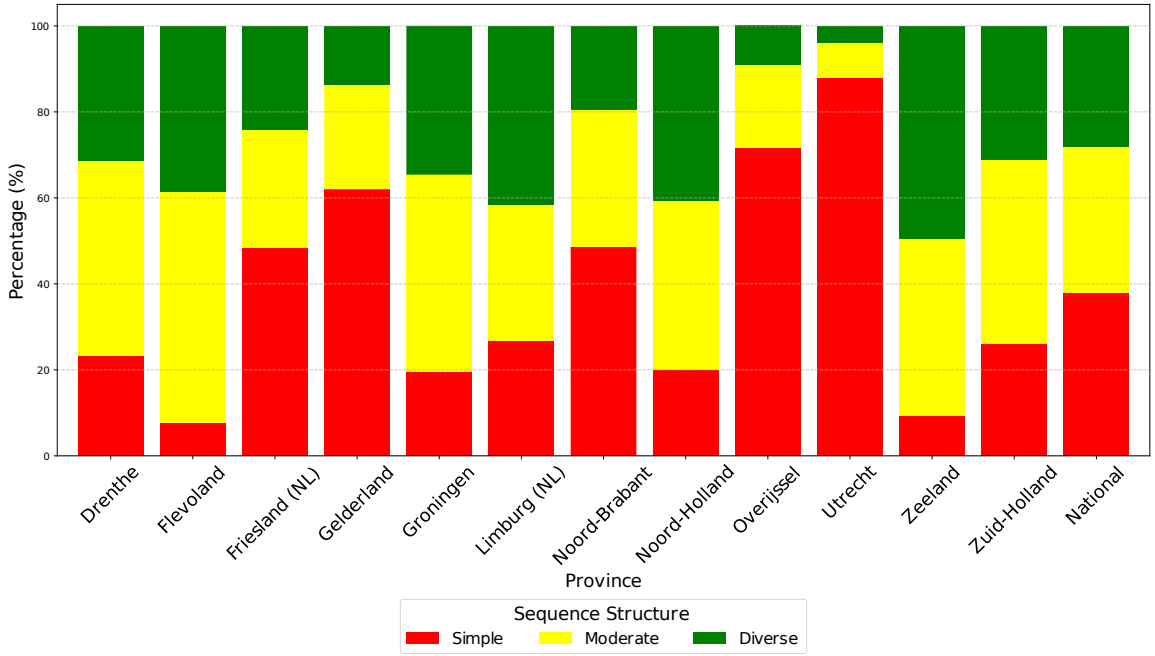


Figure 11: Main crop diversity in the percentage of the BRP per province

In Figure 13d-e-f the dominant cluster for each region is depicted, revealing a strong overlap between the soil texture shown in Figure 13c and the dominant clusters. The eastern and western regions, characterized by sandy loam soil textures, are predominantly represented by cluster 44, which corresponds to the maize monoculture (Table 1). The second most prevalent cluster in these areas is cluster 56, associated with the maize-grass rotation. An analysis of the third most dominant rotations in these areas reveals a more diverse pattern, through these clusters often involve rotations in which grass or maize are frequent crops.

The loamy soils in Flevoland, Zeeland & Noord-Holland are primarily dominated by cluster 266, which represents a crop rotation of sugar beets, potatoes and winter wheat. Similarly, in the fertile silt-loamy soils of southern Limburg, this same crop rotation is prevalent. In contrast, the northeastern part of the Netherlands, with its sandy soils, exhibits a distinct pattern on the cluster map. In this area cluster 371, which is characterized by a high frequency of potatoes in its rotation, is dominant.

A closer examination of cover crop frequencies per province reveals a clear influence of soil textures. The sandy regions, such as Noord-Brabant, Overijssel and Limburg, show a relatively high cover crop usage in Figure 12. As discussed in section *3.1.2*, a strong relationship exists between clusters where maize is rotated frequently and the use of cover crops. A comparison between provinces with higher cover crop frequencies and regions dominated by maize clusters underlines this relationship.

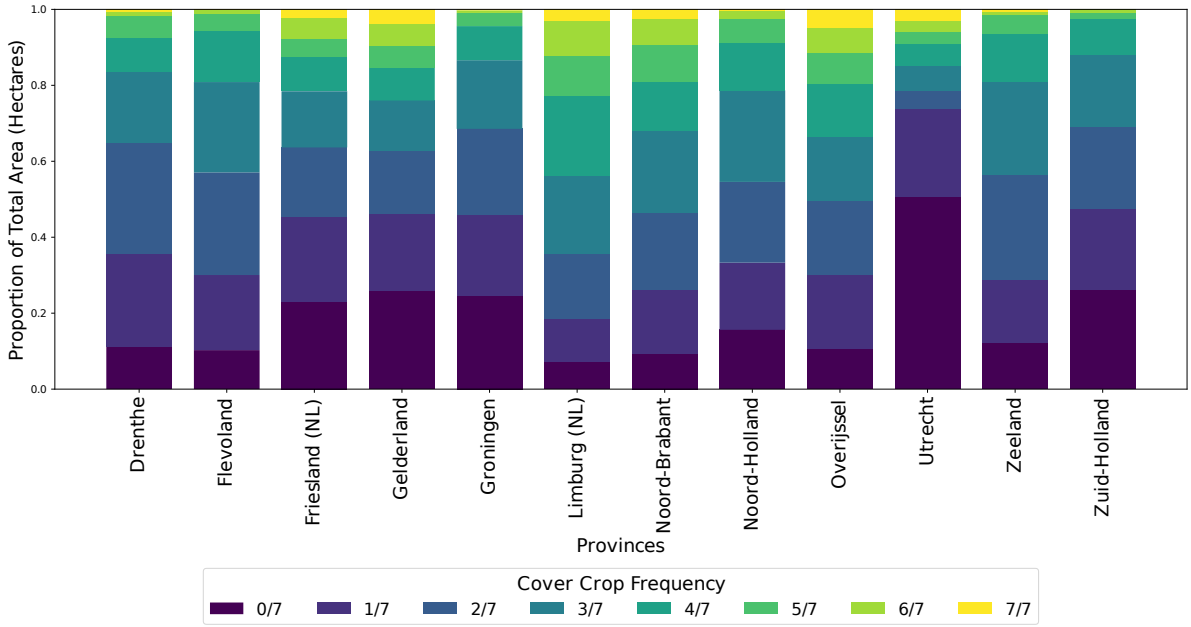


Figure 12: Cover crop usage, in the period 2017 to 2023, in hectares per province

Another interesting spatial relationship is the distribution of the most common cover crop per region, depicted in Figure 13g-h-i. The distinction between sandy-loam and loam areas is once again evident. In the western regions, leaf radish emerges as the most dominant cover crop, while in the eastern regions, the “other green manures” class prevails.

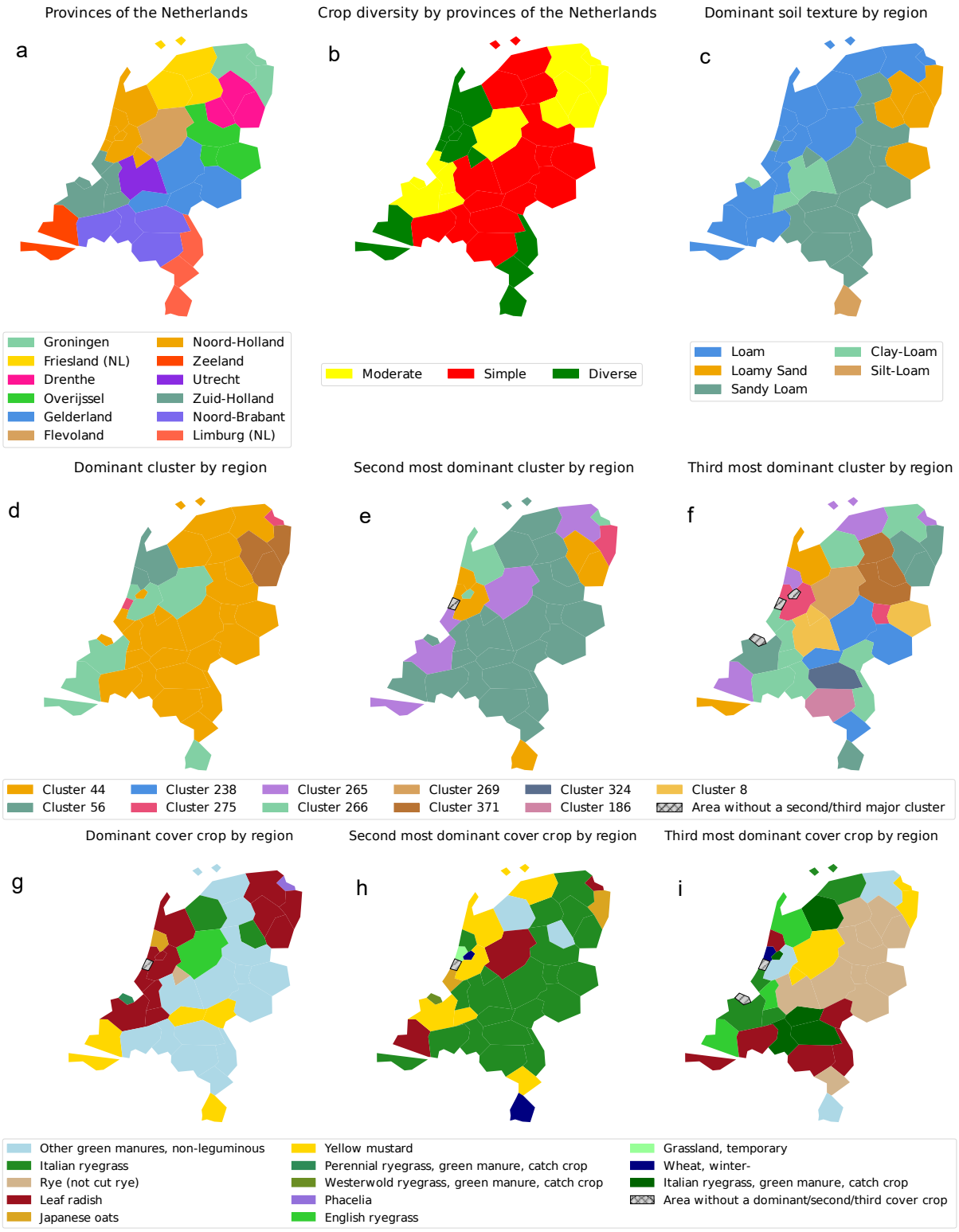


Figure 13: A collection of maps that show: Provinces (a), Crop diversity by provinces (b), Dominant soil texture by region (c), the top 3 most frequent major crop rotation clusters by region (d, e & f) and the top 3 most frequent cover crop types per region (g, h & i).

3.2 Transformer model evaluation

This paragraph shows the results of the second approach: machine learning. It will explore the potential of using the transformer model to predict future crops. To address the overarching research question – To what extent can a transformer-encoder model be used to predict future crops, while solely being trained on previous crop rotations? – this subchapter has been structured around three sub-results, each targeting a different aspect of evaluating the model: 1) showing how the model compares in terms of top-1 & top-3 accuracies, 2) showing how the model compares in prediction distribution, for the full dataset (quantitively) and 3) by zooming in on single sequences (qualitatively) by analysing its ability to reconstruct sequences using only a single crop as input.

3.2.1 Training results

After extensive trial-and-error testing, the optimum model size was identified as relatively small, consisting of 4 attention heads and 2 layers. During training the validation loss was closely monitored, and saved at its lowest value. As shown in Figure 14, this optimum was reached after approximately 5 hours of training.

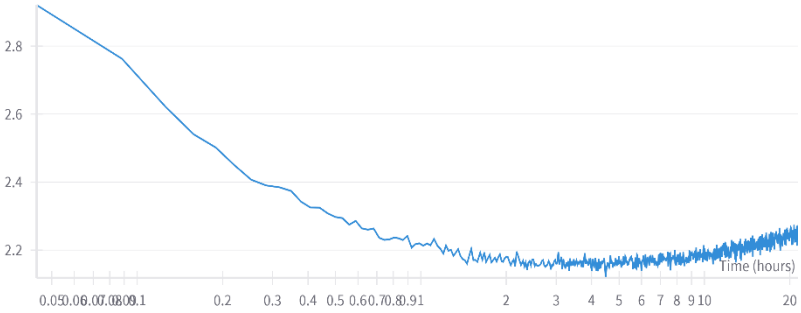


Figure 14: Training log of the validation loss for the final model

Figure 15 depicts the test metrics, including top-1 and top-3 accuracy, demonstrating consistent performance of the model (TD). Additionally, for comparative purposes, the test was evaluated using LD, CD, and ED metrics. RD is not included, as it is inherently accurate (100%) due to it being the true label. Figure 15 also shows that the transformer model outperforms both CD as well as ED, although it struggles to match the accuracy achieved by the LD. A notable trend observed is the variation in accuracy based on sequence length. The model performs best with a sequence length of 5, with minimal performance difference observed across other lengths. Interestingly, the LD outperforms the TD by a larger difference on longer sequences, while for shorter sequences this difference is much smaller.

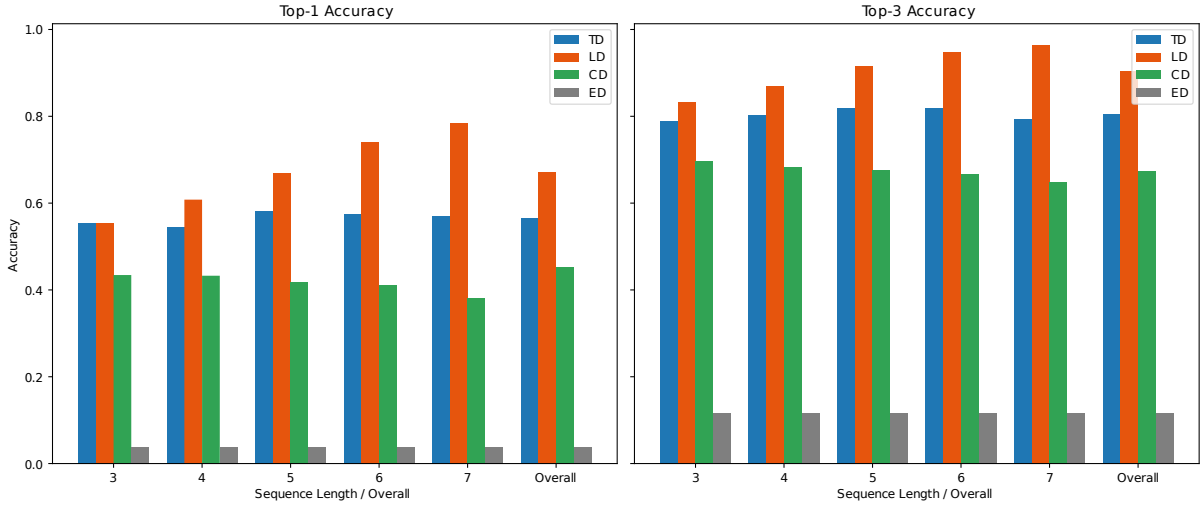


Figure 15: Top-1 and top-3 test accuracies compared over different sequence lengths and different prediction strategies.

3.2.2 Model comparison across length and frequency

The KL-divergence was computed between the four comparative distributions and the reference distribution (RD), after splitting the dataset into three distinctive frequency groups: common, uncommon, and rare. The results are depicted in Figure 16, which shows the distribution of KL-divergence values. Comparing the distributions of KL-divergences in boxplots shows how consistent the approaches compare to each other, in both median values as well as outliers, a smaller box implies a more consistent prediction accuracy than a bigger box. A key observation is the difference in predictive performance across the frequency groups, whereas the common and uncommon groups handle relatively lower KL-values, with CD/LD/TD mostly lower than five, the rare frequencies show an interquartile range (IQR) above 10 for these distributions. For the common sequences, TD consistently underperforms compared to LD, and to a lesser extent also to CD. Looking at the uncommon divergences, it can be seen that CD & LD are susceptible to a generous amount of outliers, compared to only a few within TD. The rare sequence, TD demonstrates a more stable distribution of KL-values than LD and CD, as indicated by a smaller IQR. Although LD achieves a better mean score, it exhibits a wider spread in its distribution, sometimes even predicting close to random guessing (ED).

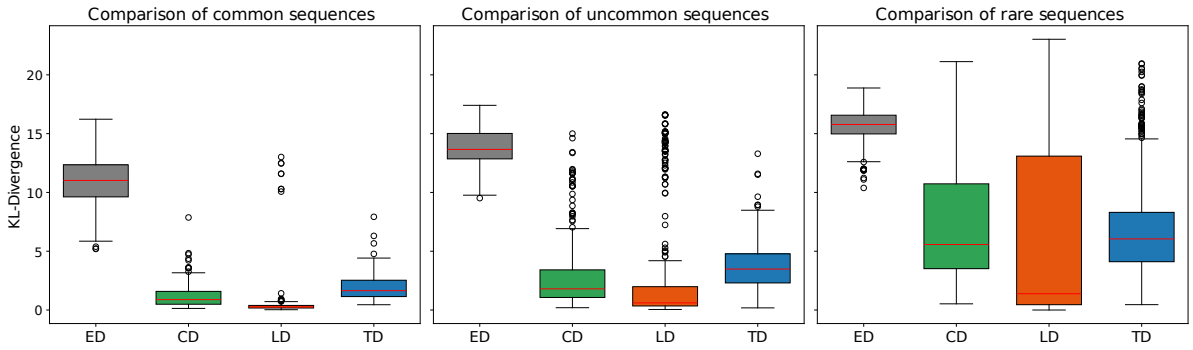


Figure 16: An overall comparison of KL-divergences values between different prediction strategies

As shown in Figure 15, the testing accuracies vary across sequence lengths, with particularly pronounced differences in the LD approach. Figure 17 presents the KL-divergence values grouped by input sequence length, with sequences of two crops displayed at the top, followed by sequences of three, four, five and six crops at the bottom. A key observation is that, in general, when sequence length increases, the KL divergence of LD decreases, whereas the predictive performance of TD remains relatively stable across the sequence lengths. For rare sequences, LD exhibits a large interquartile range, and for shorter rare sequences it is even outperformed by both TD and CD.

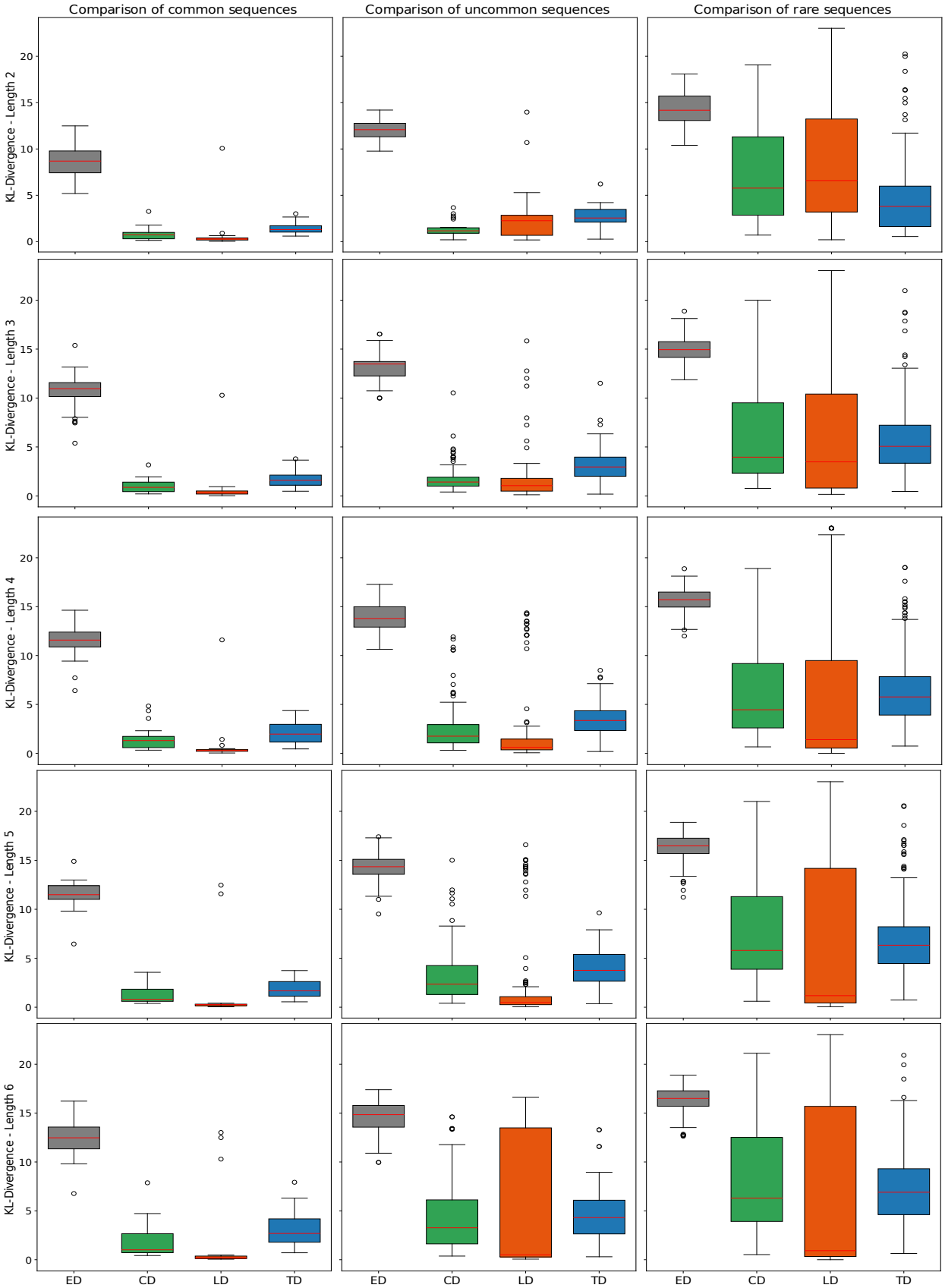


Figure 17: A comparison between distributions of KL-divergences at different sequence lengths and frequency groups

3.2.3 Prediction performance of individual sequences

To further examine the differences in prediction performance, a qualitative analysis has been conducted on individual sequences. Figure 18 illustrates an example of a rare sequence with length 3. In this case, the transformer-based approach (TD) correctly predicts potatoes as the next crop, whereas the LD approach significantly misclassifies the sequence, predicting either summer barley or sugar beets. Notably, TD assigns a relatively low probability to maize, the most commonly cultivated crop, as indicated by the CD approach.

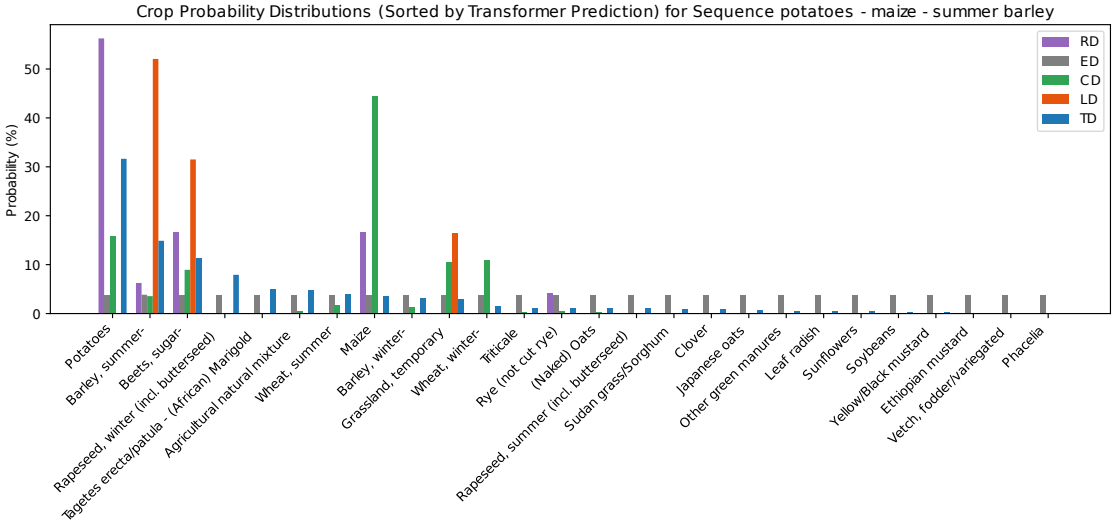


Figure 18: A qualitative analysis of the input sequence potatoes - maize - summer barley

Another example, shown in Figure 19, examines the opposite case: the most common sequence of six-year-long maize monoculture. While TD correctly predicts maize as the next crop, its prediction is relatively conservative, assigning only 40% probability to maize while distributing the remaining probability across a range of unlikely crops. This leads to a higher KL-divergence value, as the predicted distribution deviates significantly from the RD approach, which assigns over 80% probability to maize. In contrast, the LD approach predicts maize with nearly 90% certainty, resulting in a lower KL-divergence score.

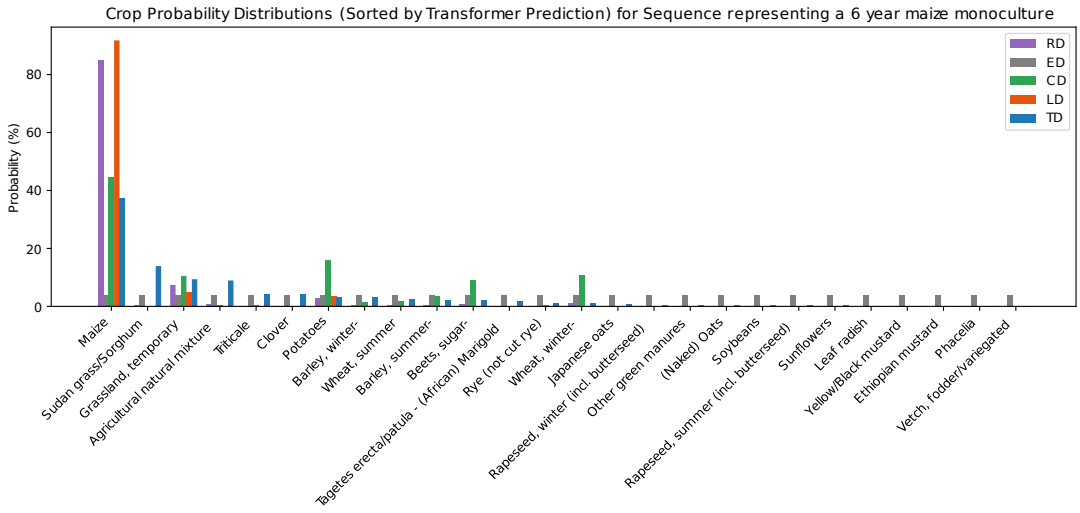


Figure 19: A qualitative analysis of the common sequence of a 6-year maize monoculture

3.2.4 Constructing sequences

The reconstruction of sequences from a single input crop yielded the metrics presented in Table 2. Among 1,000 random sequences that were used as input, the transformer model successfully reconstructed 396.

Table 2: Results of reconstructing 1000 sequences

Metric	Value
Total sequences processed	1000
Correct predictions	396
Incorrect predictions	604
Accuracy (%)	39.6

4 Discussion

4.1 Hierarchical clustering

The first part of this thesis aimed to explore the use of crop rotations, and its relationship to the use of cover crops, in the Netherlands. The clustering analysis identified 18 major crop rotation patterns, which align well with the known agricultural practices in the Netherlands. For example, cluster 44, the largest identified cluster, represents a maize monoculture, which is a widely adopted farming practice in the Netherlands (Wesselink & potters, 2022). Furthermore, several clusters feature frequent rotations of fodder crops such as maize and grass, reinforcing the validity of the cluster approach in capturing real-world agriculture systems. Another notable finding is the presence of multiple clusters incorporating potatoes every three to four years. This reflects a common strategy to mitigate the risk of yield loss due to soilborne diseases like Phytophthora infestans, which are aggravated by continuous potato cultivation (Johnson & Dung, 2010).

The results also reveal a strong relationship between soil texture and cropping diversity, as shown in Figures 13a & 13b. Regions with loamy soil exhibit higher crop diversity, which can be explained by the predominance of arable farming on these fertile soils. In contrast, regions with sandy loam soils, display a lower crop diversity, as they are primarily dominated by dairy farming, where most farmers only cultivate the fodders crops maize and grass (Aarts et al, 1999).

A note-worthy case is the province of Utrecht, which contains extensive low-lying clayey/peaty polders. These soils are stiff and harder to cultivate arable crops, leading to an agricultural landscape dominated by intensive dairy farming, as grass excels on these soils (Provincie Utrecht, 2015). Consequently, the region is characterized by rotation with a high prevalence of fodders crops such as grass and maize.

The spatial distribution of the identified clusters aligns well with the soil texture map. Cluster 44, the maize monoculture, is dominant in the sandy-loam region in the eastern and southern parts of the country, as well as around the province of Utrecht. The second and third- most dominant clusters in these areas feature high frequencies of maize and grass, underlining the strong presence of dairy farming. Conversely, in the loamy regions of the western Netherlands and Flevoland, where arable farming is more prevalent, crop rotations clusters are dominated by sugar beets, potatoes and cereals such as barley and wheat.

The findings indicate that sandy-loam areas have the highest cover crop usage over the seven-year thesis period. This trend is primarily driven by environmental policies, as sandy soils are highly susceptible to nitrogen leaching. Incentive programs promoting cover crop adoption have been in place for over two decades (Brussaard, 1992). Since 2019, regulatory enforcement has intensified, mandating that cover crops in maize cultivation on sandy soils be established before October 1st (Fan et al., 2020). One such example is maize, the dominant crop in most of the found clusters in these regions. This effect is underlined by (Kathage et al, 2022) who asked farmers in Overijssel, a province where cluster 44 is dominant, if they cultivate cover crops mandatory or voluntarily, 85.9% answered that they plant it because it is mandatory.

Coming back to the province of Utrecht, the effect of nitrogen regulation (Ministerie van Landbouw, Natuur en Voedselkwaliteit, 2021) on cover crop adoption is particularly evident. Despite being dominated by clusters with a high frequency in maize cultivation, Utrecht exhibits relatively low cover crop usage. This discrepancy can be explained by policy distinctions based on soil type. While the use of cover crops after maize is mandated on sandy-

loam soils to reduce nitrogen leaching, clayey soils, such as those in Utrecht, are less prone to this issue. Therefore, the use of cover crops is not as widely implemented. Interesting to point out is the fact that from 2024 onwards, regulations regarding cover crops and maize cultivation will also apply to designated clay soils classified as nutrient pollution areas (RVO, 2023), and thus an increase in cover use is expected in this area.

The types of cover crops used also vary by region, cluster, and soil texture. In maize-dominated areas, farmers predominantly select cost-effective cover crops such as green manures, typically a mixture of fast-growing species approved to use as cover crops, or grasses, which serve a dual purpose as fodder crops. This suggests that the choice of cover crop types is primarily driven by regulatory requirements rather than agronomic benefits, a suggestion also made by Kathage et al (2022).

In contrast, arable farming regions exhibit a different cover crop selection strategy, focusing on species that enhance crop rotations and improve soil fertility. This difference underscores the varying motivation behind choices made around cover crops across different agricultural landscapes, with regulatory compliance being a key driver in diary-dominated regions and agronomic optimization playing a more significant role in arable farming areas.

4.2 Comparing the transformer to baseline distributions

A comparison of testing accuracy (Figure 15) highlights the importance of sequence length in crop prediction when using the lookup-based (LD) approach, whereas the predictions generated by the transformer (TD) are less affected by sequence length. This suggests that TD offers an advantage by being less sensitive to the amount of input data available for a single field prediction, whereas LD performs significantly better when provided with a more extensive field history.

To further analyse these results, sequences were categorized into three frequency-based groups: *common, uncommon,* and *rare.* Generally, rare sequences proved more challenging to predict. For common sequences, the weighted crop prediction approach (CD) outperformed the model (TD), indicating that these common sequences primarily consist of the most frequently occurring crops. However, as sequence frequency decreased, the performance gap between TD matched CD in median Kullback-Leibler divergence values while exhibiting a smaller interquartile range (IQR), indicating greater robustness to outliers. A similar, albeit less pronounced, trend was observed when comparing TD to LD: while the KL divergence gap between the distributions decreased with decreasing sequence frequency, LD still outperformed TD in median values. However, the bigger IQR of LD suggests that while it excels at predicting certain distributions, it also fails on many sequences by a large margin. For uncommon sequences, the effect was less pronounced, but the presence of numerous outliers suggests that LD effectively predicts well-known cropping patterns however struggles with less common rotations, a limitation which TD does not show.

Examining the performance across different sequence lengths further supports the conclusion that TD remains stable across varying input lengths, unlike LD and CD. This can be attributed to TD's training process, which involved recognizing patterns in both short and long sequences, thereby reducing its dependence on extensive input data. The frequency-based performance trends observed in the overall evaluation also hold across different sequence lengths, with TD outperforming LD specifically on short, rare sequences.

The investigation of the prediction performance of individual sequences further reinforces the statement that TD handles rare sequences better than LD. TD correctly predicts the next crop

in the example, however, assigns a conservative probability and covers a range of small probabilities. A similar pattern emerges for the most common sequence in the dataset, the maize monoculture. While TD correctly predicts the following crop to be maize, it underestimates its likelihood by 40%, whereas LD rightly strongly favours Maize as the primary outcome. Unlike LD, TD distributes again the probabilities more broadly, assigning higher-than-expected probabilities to alternative fodder crops such as sorghum and grass. This conservative behaviour likely stems from the weighted loss function applied ring training, which was designed to prevent the model from overfitting to dominant crops in an imbalanced dataset. It also underlines the model's ability to recognize common alternatives, in this case, sorghum and grass, for crop rotations.

Although brief, the final result paragraph provides an insight into the underlying reasoning of the model. A prediction accuracy of 39.6% may appear low at first glance; however, it demonstrates the model's capacity to reconstruct sequences with only a single crop as input by recognizing and generating exact copies of major crop rotation sequences.

4.3 Limitations

4.3.1 The data science approach

The hierarchical clustering approach in this thesis is subject to several limitations that may affect the representation and interpretation of the found crop rotation patterns. Firstly, the analysis was restricted to a sequence of the exact length of seven years, which, while informative, does not capture the full spectrum of crop rotation lengths used by Dutch farmers. Some rotations may be short or longer, and excluding these variations could have led to an incomplete representation of actual farming practices.

Secondly, the clustering method employed a Hamming distance threshold of three, meaning that sequences that differ on more than 3 exact spots were not assigned to the same cluster. This threshold may have led to the misclassification of shorter, more intensive crop rotations, as even small variation, for example, a repetition of two years of winter wheat instead of one, would cause the entire sequence to be treated as significantly different. While the introduction of a shifted Hamming distance helps to mitigate sensitivity to exact starting years, the method still relies on farmers following identical crop orders. This limitation may have reduced the clustering approach's ability to recognize functionally similar rotations however includes minor variations.

Furthermore, the dataset used for clustering was a sampled subset of the full dataset of Dutch agricultural fields. While this approach ensured computational efficiency, it may have led to the underrepresentation of small regionally specialized clusters. Some of these clusters were detected but categorized as "other" due to their limited cluster size, as this thesis only aimed at analysing major crop rotation clusters.

4.3.2 The machine learning approach

The primary limitations of the transformer model lie in the scope of information it is trained on, or more specifically, the lack thereof. The model predicts the next crop in a rotation solely based on historical crop sequence, yet in real-world agricultural decision-making, crop choice is influenced by many factors that go beyond past rotations.

4.3.3 Missing contextual and external factors

One significant limitation is the absence of region-specific variables such as soil type, farm size, and proximity to advanced agricultural industries. Soil properties influence crop

suitability, as some crops grow better under specific soil conditions. This influences the crop rotation, as shown in the clustering approach. Farm size can determine the feasibility of cultivating certain crops due to machinery requirements and operation constraints (Edwards, 2017). Proximity to processing industries, for example, the ethanol plants, also plays a role in crop choice, as localized demand can drive specific cropping patterns (Park et al, 2019). Additionally, changes in land ownership can drastically alter crop rotations; for example, a field previously managed by a dairy farmer may transition from grass-based forage production to a diverse arable cropping pattern under new ownership.

Economic and policy-related factors are also unaccounted for in the model. Seed availability and price fluctuations influence crop selection, as farmers adjust planting decisions in response to cost and expected market returns. Similarly, agricultural policies, including subsidies and environmental regulation, impose constraints or incentives that shape crop choices. These external drivers are highly dynamic and are not currently integrated into the model, limiting its applicability in capturing real-world decision-making processes.

4.3.4 Data constraints and preprocessing decisions.

The dataset itself presents limitations that may affect model performance. The exclusion of non-declared data, for example, was an intentional preprocessing choice to simplify predictions. However, it could be that these non-declared fields where intentionally left fallow, which is in some farming systems a common crop rotation practice.

Another limitation stems from the temporal scope of the dataset, which spans only seven years (2017–2023). While a larger dataset would increase computational demands, it would also allow the model to capture multiple crop rotation cycles for a single field, potentially improving long-term predictive accuracy.

Furthermore, the model was trained only on the primary crop for each year (Crop1), ignoring secondary crops (Crop2) that appear in more intensive farming systems. This decision was made to simplify training and improve sequence learning efficiency. However, in regions where multiple cropping per year is common, excluding secondary crops may lead to incomplete or less accurate training sequences.

5 Conclusion and outlook

This thesis aimed to 1) detect and analyse crop rotations in the Netherlands and their relation to the use of cover crops and 2) evaluate the potential of using a transformer encoder model for predicting future crops based on crop rotation sequences.

5.1 Clustering

The 18 identified major crop rotation patterns align well with known Dutch farming practices. Especially the largest cluster, the maize monoculture, exemplifies the widespread need for fodder crops in the sandy-loam region. Secondly, the common strategy to integrate potatoes in arable crop rotations once in three to four years has correctly been clustered.

A strong correlation was observed between soil texture and crop diversity. Loamy soil regions exhibited higher crop diversity, supporting intensive arable rotations, whereas sandy-loam regions had lower diversity, dominated by fodder crops such as maize and grass. This pattern also extends to cover crop adoption, with sandy-loam areas displaying the highest cover crop usage. This is mostly explained by environmental policies aimed at reducing nitrogen leaching in these regions.

The choice of cover crop species varied significantly by region and dominant crop rotation. In the sandy-loam maize-dominated areas, cover crop selection is primarily driven by compliance with environmental regulations rather than agronomic benefits. Conversely, arable farming regions showed a dominant appearance of crops that enhance crop rotations. These findings highlight the regulatory and agronomic motivations behind cover crop adoption across different agricultural landscapes.

5.2 Transformer model

The transformer-encoder model demonstrated that while mostly underperforming in comparison to the look-up approach, it can acceptably predict the next crop. The model particularly showed its potential to handle infrequent crop rotations, outperforming the other methods in scenarios with limited input data.

Furthermore, a key finding was the stability in accuracy that the model could provide across different input sequence lengths. While the non-deep learning approach LD performed exceptionally well for common sequences, it exhibited high variability in successful predictions and struggled with less frequent crop rotations. The transformed model, in contrast, showed a consistent prediction pattern, demonstrating its adaptability to a diverse range of sequences.

The investigation of the performance of individual sequences revealed that the model's underperformance is partly explained by its conservative approach in assigning a probability distribution. This conservative approach may create a higher KL-score, however looking at individual sequences this is explained by the fact that the models does not only capture the dominant rotations. It also recognizes plausible alternatives, reinforcing the statement that the model has the potential to be used to simulate farmer decisions. The model showed the ability to reconstruct a part of the sequences with a little input, underlining that it was able to learn common crop rotation patterns.

5.3 Limitations and future directions

Both approaches have limitations. The clustering method, restricted to a seven-year sequence length, may have overlooked the extensive variation found in real-world crop rotation lengths. Additionally, the imposed Hamming distance threshold may have misclassified functionally similar rotation with minor sequence variations. Future research is advised to refine the clustering techniques and incorporate flexible sequence lengths.

The primary limitation of the transformer model is its reliance on solely historical crop sequences, without incorporating often known regional variables, such as soil type or economic factors. These variables play a crucial role in farmers' decision-making, and integrating such data could significantly improve model performance. To reach a fully functioning, and trustable AI tool to predict farmers' decision-making, future work should incorporate these sequences.

5.4 Final remarks

Overall, this thesis has provided a comprehensive understanding of Dutch major crop rotations and their relations to several key variables like soil texture, diversity & cover crop use. The data science approach was successfully put to use in the creation & evaluation of a transformer-encoder model to predict future crops. The clustering results have highlighted the strong influence of soil texture and environmental policies on crop rotation practices. Meanwhile, the transformer model demonstrated that the extensive BRP dataset on cropping history is a promising basis for creating a predictive tool for agricultural planning, particularly the prediction of less common crop rotations. Despite the mentioned limitations, this thesis marks a step forward in the use of a data-driven approach to better simulate and understand the effect of agricultural policies, paving the way for future research on decision-making AI support systems.

6 Use of generative AI statement

AI, in the form of OpenAI’s ChatGPT, has been used as a sparring partner to 1) act as a coding assistant. In this role it provided debugging advice, improved code efficiency & generated basic code examples: 2) To check grammar & spelling and improve fluency in reading

7 References

Cover Image: suedzucker.com (2025). *Sugar Beets as a Central Crop in Crop Rotation Fostering Biodiversity and Soil Health. Retrieved from* https://www.suedzucker.com/sugar-beets-as-a-central-crop-in-crop-rotation-fostering-biodiversity-and-soil-health/

3e Nederlandse Actieprogramma (2004-2009) inzake de Nitraatrichtlijn; 91/676/EEG. (2005). In [S.l.], NL, Ministerie van Landbouw, Natuur en Voedselkwaliteit. http://edepot.wur.nl/117897

5e Nederlandse AP betreffende de Nitraatrichtlijn (2014-2017). (2014). In [S.l.], NL, Ministerie van Landbouw, Natuur en Voedselkwaliteit. https://edepot.wur.nl/342467

6e Nederlandse actieprogramma betreffende de Nitraatrichtlijn (2018 - 2021). (2017). In [S.l.], NL, Ministerie van Landbouw, Natuur en Voedselkwaliteit. https://edepot.wur.nl/425609

7e Nederlandse actieprogramma betreffende de Nitraatrichtlijn (2022-2025). (2021). In Den Haag, Ministerie van Landbouw, Natuur en Voedselkwaliteit [etc.]. https://edepot.wur.nl/558225

Aarts, H. F. M., Habekotté, B., Hilhorst, G. J., Koskamp, G. J., van der Schans, F. C., & de Vries, C. K. (1999). *Efficient resource management in dairy farming on sandy soil*. Research Institute for Agrobiology and Soil Fertility (AB).

Ballot, Rémy & Guilpart, Nicolas & Jeuffroy, Marie-Helene. (2023). *The first map of crop sequence types in Europe over 2012–2018*. Earth System Science Data. *15. 5651-5666. 10.5194/essd-15-5651-2023.*

Bengio, Y. & Simard, Patrice & Frasconi, Paolo. (1994). Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council. 5. 157-66. 10.1109/72.279181.

Bi, L., Owen, W., Guiping, H., Tenuta, A. U., Kandel, Y. R., & Mueller, D. S. (2023). A transformer-based approach for early prediction of soybean yield using time-series images. Frontiers in Plant Science, 14, 1173036. https://doi.org/10.3389/fpls.2023.1173036

Al Kindhi., Muhammad, Afif, Hendrawan., Diana, Purwitasari., Tri, Arief, Sardjono., Mauridhi, Hery, Purnomo. (2017). Distance-based pattern matching of DNA sequences for evaluating primary mutation. doi: 10.1109/ICITISEE.2017.8285518

Brussaard, W. (1992). Agrarian land law in the Netherlands. *Agrarian land law in the western world*, 114-133.

Edwards, W. (2017). *Farm machinery selection*. Iowa State University, Ag Decision Maker. Retrieved from https://www.extension.iastate.edu/agdm/crops/pdf/a3-28.pdf

European Union. (1991). Council Directive 91/676/EEC of 12 December 1991 concerning the protection of waters against pollution caused by nitrates from agricultural sources. *Official Journal of the European Communities, L375, 1–8.* Retrieved from EUR-LEX.EUROPA.EU

R. W. Hamming, "Error detecting and error correcting codes," in The Bell System Technical Journal, vol. 29, no. 2, pp. 147-160, April 1950, doi: 10.1002/j.1538-7305.1950.tb00463.x.

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for Clustering Data.* Prentice-Hall.

Johnson, D. A., & Dung, J. K. S. (2010). Rotation and cover crop effects on soilborne potato diseases, tuber yield, and soil microbial communities. *Plant Disease*, *94*(12), 1491–1502. https://doi.org/10.1094/PDIS-03-10-0172

Ndikumana, E., Ho Tong Minh, D., Baghdadi, N., Courault, D., & Hossard, L. (2018). Deep Recurrent Neural Network for Agricultural Classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sensing*, *10*(8), 1217. https://doi.org/10.3390/rs10081217

Kathage, J., Smit, B., Janssens, B., Haagsma, W., & Adrados, J. L. (2022). How much is policy driving the adoption of cover crops? Evidence from four EU regions. *Land Use Policy*, *116*. https://doi.org/10.1016/j.landusepol.2022.106016

Kaspar, T.C. and Singer, J.W. (2011). The Use of Cover Crops to Manage Soil. In Soil Management: Building a Stable Base for Agriculture (eds J.L. Hatfield and T.J. Sauer). https://doi.org/10.2136/2011.soilmanagement.c21

Kullback, S., & Leibler, R. A. (1951). *On information and sufficiency.* Annals of Mathematical Statistics 22(1), 79-86. 10.1214/aoms/1177729694

Lankoski, J., & Thiem, A. (2020). Linkages between agricultural policies, productivity and environmental sustainability. *Ecological Economics, 178*, 106809. https://doi.org/10.1016/j.ecolecon.2020.106809

LeCun, Yann & Bengio, Y. & Hinton, Geoffrey. (2015). Deep Learning. Nature. 521. 436-44. 10.1038/nature14539.

Liu, X., Beusen, A. H., Van Grinsven, H. J., Wang, J., Van Hoek, W. J., Ran, X., Mogollón, J. M., & Bouwman, A. F. (2024). Impact of groundwater nitrogen legacy on water quality. *Nature Sustainability*, *7*(7), 891-900. https://doi.org/10.1038/s41893-024-01369-9

Ofitserov, E., Tsvetkov, V., & Nazarov, V. (2019). *Soft edit distance for differentiable comparison of symbolic sequences*. http://arxiv.org/abs/1904.12562

Park, J., Anderson, J., & Thompson, E. (2019). Land-Use, Crop Choice, and Proximity to Ethanol Plants. *Land*, *8*(8), 118. https://doi.org/10.3390/land8080118

Pelletier, C., Webb, G. I., & Petitjean, F. (2019). Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, *11*(5). https://doi.org/10.3390/rs11050523

Provincie Utrecht. (2015.). *Factsheet 12: Rivierengebied – rivierklei, vochtig tot nat*. Provincie Utrecht. https://geo.provincie-utrecht.nl/publiek/documenten/bodem/bodemsysteemwijzer/factsheets/12.pdf

RVO - Rijksdienst voor Ondernemend Nederland. (2023.). *Vanggewas na mais op klei- en veengrond in NV-gebieden*. Retrieved march 10, 2025, from https://www.rvo.nl/onderwerpen/mest/vanggewas/vanggewas-na-mais-op-klei-en-veengrond

Schils, R. L. M., Aarts, H. F. M., Bussink, D. W., Conijn, J. G., Corré, W. J., van Dam, A. M., Hoving, I. E., van der Meer, H. G., & Velthof, G. L. (2002). Grassland renovation in the Netherlands; agronomic, environmental and economic issues. In J. G. Conijn, G. L. Velthof, & F. Taube (Eds.), *Grassland resowing and grass-arable crop rotations; international workshop on agricultural and environmental issues, Wageningen, the Netherlands, 18 & 19 April 2002* (pp. 9–24). Plant Research International.

Stein, S., & Steinmann, H. H. (2018). Identifying crop rotation practice by the typification of crop sequence patterns for arable farming systems – A case study from Central Europe. *European Journal of Agronomy*, *92*, 30–40. https://doi.org/10.1016/j.eja.2017.09.010

Ashish, Vaswani., Noam, Shazeer., Niki, Parmar., Jakob, Uszkoreit., Llion, Jones., Aidan, N., Gomez., Lukasz, Kaiser., Illia, Polosukhin. (2017). Attention is All you Need. 30:5998-6008.

Velthof, G. L., van Schooten, H. A., & van Dijk, W. (2020). Optimization of the nutrient management of silage maize cropping systems in The Netherlands: A review. *Agronomy, 10*(12), Article 1861. https://doi.org/10.3390/agronomy10121861

Wen, Qingsong & Zhou, Tian & Zhang, Chaoli & Chen, Weiqi & Ma, Ziqing & Yan, Junchi & Sun, Liang. (2022). Transformers in Time Series: A Survey. 10.48550/arXiv.2202.07125.

Wesselink, M., & Potters, J. (2022, June 28). *Case study 1: The Netherlands – Breaking maize monoculture*. Wageningen University & Research. Retrieved from https://www.diverimpacts.net/case-studies/case-study-1-nl.html

Wang, P., Xie, W., Ding, L., Zhuo, Y., Gao, Y., Li, J., & Zhao, L. (2023). Effects of Maize–Crop Rotation on Soil Physicochemical Properties, Enzyme Activities, Microbial Biomass and Microbial Community Structure in Southwest China. *Microorganisms*, *11*(11). https://doi.org/10.3390/microorganisms11112621

8 Annexes

List of research data to be submitted after examination:

- The complete field parcel data set
- Github, containing cod, processed datasets, graphs and figures

8.1 Annex I: Metrics

	Length 3	**Length 4**	**Length 5**	**Length 6**	**Length 7**	**Overall score**
TD Top-1 Accuracy	0.553219	0.545121	0.581214	0.574226	0.569680	0.5646
Top-3 Accuracy	0.788765	0.802767	0.819734	0.818182	0.794185	0.8047
Top-1 LD Accuracy	0.5542	0.6080	0.6682	0.7409	0.7835	0.6709
Top-3 LD Accuracy	0.8323	0.8699	0.9155	0.9478	0.9650	0.9052
Top-1 CD Accuracy	0.434297	0.432725	0.418456	0.411662	0.380807	0.4516
Top-3 CD Accuracy	0.697438	0.683454	0.675917	0.665901	0.647612	0.674076
ED Accuracy	0.038462	0.038462	0.038462	0.038462	0.038462	0.038462

	Length 3	**Length 4**	**Length 5**	**Length 6**	**Length 7**	**Overall score**
Top-1 Precision	0.590095	0.584220	0.619566	0.620172	0.614711	0.6046
Top-1 F1-score	0.558654	0.549610	0.593628	0.591786	0.586658	0.5778
Top-1 Recall	0.553219	0.545121	0.581214	0.574226	0.569680	0.5646