UiA University of Agder

Master

Hybrid Neural Networks with Attention-based Multiple Instance Learning for Improved Grain Identification and Grain Yield Predictions

MIKKEL ANDREAS KVANDE
SIGURD LØITE JACOBSEN

SUPERVISOR
Rashmi Gupta and Morten Goodwin

University of Agder, 2022
Faculty of Engineering and Science
Department of Engineering and Sciences

Acknowledgments

We want to thank the Faculty of Engineering and Science and the CAIR Research lab at the University of Agder, Norway, for allowing us to conduct this research effort. We thank Felleskjøpet Agri, NLR, and inFuture, who coordinated the KORNMO project– production optimization, quality management, and sustainability across the grain value chain. KORNMO provided much of the data, while Simon Arenberg at inFuture, offered valuable insight into the existing data and previous research efforts.

Special thanks are given to our supervisors, Dr. Rashmi Gupta and Prof. Dr. Morten Goodwin, for their great and consistent supervision throughout the entirety of our master's thesis.

Abstract

Agriculture is a critical part of the world's food production, being a vital aspect of all societies. Procedures need to be adjusted to their specific environment because of their climate and field condition disparity. Existing research has demonstrated the potential of grain yield predictions on Norwegian farms. However, this research is limited to regional analytics, which is unable to acquire sufficient plant growth factors influenced by field conditions and farmers' decisions. One factor critical for yield prediction is the crop type planted on a per-field basis.

This research effort proposes a novel approach for improving crop yield predictions using a hybrid deep neural network utilizing temporal satellite imagery from a remote sensing system. Additionally, We apply a variety of data, including grain production, meteorological data, and geographical data. The crop yield prediction system is supported by a field-based crop type classification model, which supplies features related to crop type and field area. Our crop classification system takes advantage of both raw satellite images as well as carefully chosen vegetation indices. Further, we propose a multi-class attention-based deep multiple instance learning model to utilize semi-labeled datasets, fully benefiting Norwegian data acquisition.

Our best crop classification model, which consists of a time distributed network and a gated recurrent unit, classifies crop types with an accuracy of 70% and is currently state-of-the-art for country-wide crop type mapping in Norway. Lastly, our yield prediction system enables realistic in-season early predictions that could benefit actors in real-life scenarios.

Contents

List of Figures

List of Tables

Abbreviations

AI Artificial Intelligence

ANN Artificial Neural Network

CNN Convolutional Neural Network

DNN Deep Neural Network

Evi Enhanced Vegetation Index

GARI Green Atmospherically Resistant Index

GNDVI Green Normalized Difference Vegetation Index

GRNDVI Green-Red Normalized Difference Vegetation Index

GRU Gated Recurrent Units, type of RNN

kg/daa Kilograms per Decare, the area-adjusted unit for crop yield used in the research effort.

LSTM Long Short-Term Memory

LSWI Land Surface Water Index

MAE Mean Absolute Error

MAPE Mean Absolute Percentage Error

MIL Mean Multiple Instance Learning

NDRE Normalized Difference Red-Edge

NDSVI Normalized Difference Senescent Vegetation Index

NDTI Normalized Difference Tillage Index

NDVI Normalized Difference Vegetation Index

NIBIO Norwegian Institute for Bioeconomy Research

RMSE Root Mean Square Error

PVR Photosynthetic Vigor Ratio

RPCA Robust Principal Component Analysis

SCCCI Simplified Canopy Chlorophyll Content Index

SIPI3 Structure Intensive Pigment Index 3

SIWSI Shortwave Infrared Water Stress Index

RNN Recurrent Neural Network

Chapter 1

Introduction

This chapter introduces the research topic, problem statement, goals, and how these goals will be achieved. Agriculture is a big part of the world's food production, and improvement in this field can substantially impact the quality and amount of global food production. One field that has shown positive results related to these needs is Artificial Intelligence (AI), as it will be shown in Section 1.1, especially prediction and classification models. This research study is supported by the importance of providing agricultural monitoring systems and advisement for farmers and government officials, which will be explained in Section 1.2. This can enable farmers to improve their work efficiency, quality of yield, and overall yield production while also enabling government officials to make better country-wide decisions. To reach these goals, a set of hypotheses will be presented in Section 1.3, covering research efforts on improved grain yield predictions and grain classification. During these efforts, we faced a number of challenges, which will be explained in Section 1.4. The major challenges were gathering new satellite images and the lack of agriculture data relevant to plant growth. As a result of the research efforts and overcoming these challenges, a set of contributions to the research community will be presented in Section 1.5. Lastly, Section 1.6 presents the structure of this thesis and the contents of all chapters.

1.1 Introduction

Agriculture is an essential part of the world's food production, being a vital factor in all societies. It has also been a crucial factor in overcoming two of the United Nation's sustainable development goals: ending world hunger and ensuring sustainable consumption and production [1]. On the contrary, agriculture is one of the worst contributors to environmental pressure. It is the primary source of toxins released into the environment, including insecticides [2]. The 2011 UNEP Green Economy report stated that agricultural operations produced 13 percent of global gas emissions [3]. Agriculture techniques need to be adapted to their specific location and environment to conserve soil, limit product waste, and manage pollution such as nitrogen and phosphorous [4]. While working on overcoming challenges within agriculture, the world population is expected to increase, and the arable farmland is expected to decrease worldwide and in Norway. This change will increase the food requirment while decrease the available production area [5, 6, 7]. Drastic changes are needed to achieve sustainable production that fulfills the consumption needs. Therefore, every resource and production in Norwegian agriculture will need to be utilized as much as possible. This includes producing as many grains as possible on a field with minimal loss and the highest quality [8]. Therefore, research within agriculture is necessary today and vital if the world is ever to achieve a sustainable world [5, 6, 7].

It has become crucial to improve and better utilize agriculture to supply food for more people over a smaller production area. As a result, research within agriculture has become more critical over the

years, and much is still needed to reach these goals [6, 7]. Because of agriculture's critical place in global food production, even simple changes resulting from research can significantly impact overall production [9]. Applying AI to agriculture for monitoring, predictions, classification, and knowledge-based assistance can help the world optimize agriculture. Reaching this goal can improve crucial situations, especially in developing countries [5, 6, 7]. Different CNNs architectures have been able to predict crop yields. Sharma et al. [10], You et al. [5], and Russello & Shan [11] applied temporal remote sensing images to CNNs, resulting in an accurate crop yield prediction. In addition, some of the researchers applied textual information to LSTMs for additional data, improving the original predictions. Early in-season predictions have also shown to be possible, where Engen et al. [12, 13] experienced an error rate increase of only 7.66% and 20.89% when predicting 13 and 18 weeks earlier, respectively. In addition, classification models have proven to be able to map crop types. Kussul et al. [14] and Ji et al. [15] applied temporal remote sensing images to multi-dimensional CNNs, which were able to extract the spatial context information from the images. The former reached an accuracy of 94.6%. Machine learning has also been shown to classify crop types using derived vegetation indices. Foerster et al. [16] used the NDVI vegetation index to classify 12 crop types, reaching accuracies between 14.5% - 97.3% between the different types. Deep learning in this field has been shown as a promising technique with the potential of strengthening precision agriculture and increasing food production and has thereby grown in popularity [9].

Previous work has demonstrated the potential of grain yield predictions on Norwegian farms, which has not yet achieved a satisfactory level. The major reason is that researchers lack measurements and farming activity data from farms for the yield predictor to achieve high precision [12, 13]. In addition, early yield prediction during the growing season can not realistically be implemented, as it is currently dependent on features made available after grain delivery. There has also been much previous work on classifying crop types based on several different data sources, which has achieved promising results [16, 17, 18]. Performing crop classification in general in this environment is innovative, as the data available is unique compared to other regions. In addition, due to Norway's location and climate, the number of crop types suitable for production is limited to only a few types. 85% of its total 3 288 497 decares of open agriculture fields in Norway are used for grain production alone, mainly consisting of barley, wheat, oats, and types of rye [19]. This is different from what previous studies are based on, seeing as they distinguish crop types from each other and not sub-types of grain [9, 20].

This research effort's motivation is a result of the improved state of Norwegian agriculture in terms of sustainability and productivity, which we present in the next section. There is a growing interest and need for technological advances in machine learning. Our work begins with analyzing previous work and the available Norwegian agriculture data. The findings determine different ways of applying this data to machine learning to improve agriculture productivity, quality, and accuracy. Expanding the work by Engen et al. [12, 13], new features and data samples were introduced, improving the accuracy for predicting grain yields on Norwegian farms compared to that of state-of-the-art [12, 13]. The data was also applied to a time-distributed CNN and a semi-supervised learning algorithm for crop type mapping to further label our field boundary data. The classification-based research effort enabled early yield predictions during the growing season by using the classification model's output to estimate the growing area for each crop type. The early yield prediction achieved competitive results to that of the state-of-the-art [12, 13], while only using features available at the prediction time.

1.2 Research Motivation

Consider supplying a farmer with a software tool that would enable them to track the growth of their crops throughout the growing season as a technique for monitoring. A software tool could be used to learn and adapt to the fields and their environment to improve grain growth. Farmers with a software tool like this can make better farming decisions such as what and when to sow on

each field. In addition, when to perform specific farming activities to maximize production and quality. A software tool could also contribute to other actors like government agencies, such as the department of agriculture. The department, actor or agency responsible for the country's food production would benefit from knowing the quantity of grain production the country will produce. Knowing this could prepare them to make better financial decisions concerning grain and other food imports and export [6, 7, 21].

The only publicly available work on yield prediction for Norwegian farms has been done by Engen et al. [12, 13]. While the researchers could predict grain yield successfully, they did so by only training their model with data from the three last growing seasons. The researchers also used data that only directly represented two growth factors, temperature, and precipitation. There are, therefore, multiple ways the training data could be improved by either adding new data samples or features that represent other essential growth factors. Improving the training data could also improve the predictor and further strengthen Norwegian agriculture's position on crop monitoring.

Many of the reasons for attempting to predict crop yields is to be able to know this information during the growing season before the crops are harvested and delivered. For Norwegian agriculture to benefit from yield prediction, a system that can accurately predict the yield early in the season while only using the available data needs to be developed. Engen et al. [12, 13] performed early yield predictions without losing too much performance, but their system was based on knowing the farmer's growing area for each crop type. This information is currently only available long after the harvest and should not be used in early yield predictions. On the contrary, each crop type's growing area and size are vital for grain yield predictions to have actual value for actors such as farmers and government agencies. This is because the quantity is not significantly useful without knowing what type of crop it represents. This problem can be solved in different ways, such as enabling farmers to notify what they have sown on each field or classifying fields between the available crop types and thereby calculating the growing area [10].

Successful crop classification enables researchers and government agencies to map what type of crop has been planted in each field for the current and past growing seasons. While this information is helpful alone, it has also been shown to contribute to important and often required information for yield prediction to be successful [10]. In addition, it can also contribute to identifying the most optimal geographical properties and soil quality for growing each type of crop when combined with crop yields [6, 7]. While crop classification has been applied to many environments, as presented in Section 1.1, there are still limited available research efforts on classification in an environment limited to only grain crops. There are also, to the best of our knowledge, no records of a general crop classification for Norwegian agriculture, with Digifarm [22] being the only one with some level of relevance.

Norwegian agriculture suffers from zero documentation regarding what is planted in specific fields. The only currently available information related to this is the type and amount of grains delivered by the farmers. Using the grain delivery information, fields belonging to a farmer who only delivered one type of grain can be labeled, resulting in a semi-labeled dataset where many fields are missing their true label. This also further strengthens the need for a grain classification system, as the information a system like this can provide does not even exist today.

These are all potential systems and solutions for applying AI and Data Science to agriculture. Improving state-of-the-art techniques for yield prediction can provide positive contributions to the agriculture community. However, this does not change the fact that grain yield prediction techniques are not yet suitable for implementation in Norway. In order to enable farmers to track the growth of their crops, one needs to develop a realistic system for early yield predictions by only using features available at the time of prediction. Early yield predictions are challenging to successfully deploy without knowing the crop type of the predicted yield. Therefore, grain type classification and mapping can provide valuable information for researchers and the agriculture community and support and enable early yield predictions. Therefore, it is in all of these actors' interest to expand

the work done by Engen et al. [12, 13] to improve the accuracy and create a system for crop type mapping to enable realistic early yield predictions.

1.3 Scope of the Project

1.3.1 KORNMO

This thesis is part of a project called KORNMO. KORNMO is a research project collaboration between Felleskjøpet[1], University of Agder (UiA)[2], and inFuture[3]. The project has the primary goal of optimizing agriculture, production, quality, and environmental footprint by combining agronomy with machine learning. The project is split into the three following parts:

- **Production Optimization** - Creating services that increase value for farmers by improving the management and prediction of data to give more accurate advice on different production choices.
- **Quality Management** - Creating services that increase value for grain customers by documenting methods used during the growing season and the quality of grains. This section involves controlling production and grain delivery and applying prediction for crop yield delivery.
- **Sustainability** - Creating data for country-wide management and key information for consumers. It includes providing advice to producers related to sustainability and information regarding environmental consequences made by agriculture.

UiA's role in KORNMO is to explore data already collected and processed by other research collaborators in the project and further process this as needed. Currently, UiA and inFuture have collected a large amount of data related to satellite images, soil, meteorology, grain production, farmers, and properties. Their goal is to apply this data to machine learning techniques further to improve existing processes concerning the project goals. One of UiA's earlier works on behalf of the KORNMO project was a group of students working on crop yield prediction for their Masters' thesis in 2021 [12, 13], which will be introduced in Chapter 3 later in this thesis.

1.3.2 Hypotheses

This research effort will take what has previously been done by Engen et al. [12, 13] into careful consideration. While they could predict grain yields with minimal loss, their system lacked the ability to perform realistic in-season yield prediction by only using data available at that time. Therefore, this research will explore additional available data and previous research efforts related to precision agriculture to design an accurate, realistic system for in-season early yield prediction and a required crop classification technique. These are both essential contributions to the project goals.

Hypothesis 1: Features related to sunlight, growth temperature, ground state, and soil quality can be applied to prediction models for more accurate grain yield predictions.

The recent work by Engen et al. [12, 13] on prediction models for Norwegian farms have successfully

[1] https://www.felleskjopet.no/
[2] https://cair.uia.no/
[3] https://infuture.no/en/

predicted grain yields, but this was done with limited environmental features. Exploring additional features could improve the predictor of the state-of-the-art yield prediction in Norway. New features will be explored and pre-processed in Section 5.1 in the same format as the daily meteorological features in [12, 13]. The deep learning models used in [12, 13] will be recreated and trained with these new features. The results of this extension will be presented in Chapter 6 6 together with a comparative analysis with the original grain yield prediction results from [12, 13]. This work will extend into the second hypothesis, where the same model will be used with data from an additional year i.e., additional growing season.

Hypothesis 2: Extending agricultural data for predicting grain yields with an additional year of data samples shows an increase in accuracy and the need to collect data over time.

Engen et al. [12, 13] concluded that the limited number of seasons with data coverage was one of the more impactful limitations concerning their models' performance. Inspired by the first hypothesis, taking a different angle by including new data samples instead of introducing new features, could improve the predictor and the situation of crop monitoring in Norwegian agriculture. Agriculture and meteorological data from the 2020 season will be retrieved and processed in Section 5.1 in the same format as the features from the previous seasons in [12, 13]. This will then be merged with the previous data and applied to the already re-implemented models. The loss and accuracy of the model will be presented in Section 6.1 and compared with the original results from [12, 13].

Hypothesis 3: Satellite images and vegetation indices can be applied to convolutional neural networks for accurate grain classification on fields.

The motivation of this research effort presented numerous reasons for attempting crop classification and mapping for Norwegian agriculture. This included the ability to map for what purpose fields are being used and perform accurate yield predictions early in the growing season. After researching different classification works, features applied, and performance achieved in Chapter 3, a time-distributed CNN architecture will be implemented in Section 5.2. Raw satellite images and derived vegetation indices will be applied to the model. Ground truth class labels of fields will also be created where possible. After training the models, Section 6.2 will present the results and compare the different methods.

Hypothesis 4: Multi-class Attention-based Deep Multiple Instance Learning is able to utilize the whole field dataset and thereby increase crop classification accuracy.

Another motivation for crop classification is that there are currently no records of what is planted and grown on each field. The only available information can, at best, result in a semi-labeled dataset. A dataset with missing information as labels might require special techniques related to semi-supervised learning methods to classify fields in Norwegian agriculture. In order to utilize the whole dataset, a semi-supervised learning method, Multiple Instance Learning (MIL) method, will be implemented in Chapter 5 to attempt to outperform a CNN-based approach, showing a comparison in Section 6.2.4.

Hypothesis 5: In-season early yield predictions are as successful when using predicted crop types limited to the data available at that time.

Throughout the motivations in Section 1.2, the importance of in-season yield prediction was ex-

plained. For Norwegian agriculture to estimate and predict the future yield, experiments need to be adapted to only use information available at the prediction time. This research effort will combine everything learned and achieved through the previous hypothesis into an early yield prediction system in Section 5.3. The crop type of every field will be predicted using only the earliest satellite images, and then farm-specific masks for each crop type will be created. This will enable the yield model to predict early in the season correctly. Its performance compared to previous early yield predictions will be shown in Section 6.3.

These hypotheses are all specific scenarios that have not been explored previously, given the environment and available data. It is our belief that these research efforts will contribute valuable information and techniques to Norwegian agriculture, food production, and research community that are similar to the goals of the KORNMO project. The hypotheses also revolve around data that is already automatically collected for each relevant day or year, which means the work can be adapted to more data as each growing season passes.

1.4 Research Challenges

This section presents significant challenges related to the research of the hypothesis presented in Section 1.3 above and where these will be addressed. This research effort revolves around a rich amount of data of different types and from different sources. Applying this data while developing innovative techniques, one will likely encounter challenges.

1.4.1 Acquiring Satellite Images

The second hypothesis requires us to acquire satellite images from the newest growing season. These images must retain a certain standard and similarity to the satellite images used for the earlier growing seasons. Therefore, Sentinel-2A images are required. While the Sentinel satellite system is open-access, it is not necessarily straightforward to acquire a set of 12-band images for each farm in Norway. While the actor such as Sentinel Hub[4] has made this easy by creating an API system that handles requests, which are often locked behind a subscription fee. This challenge is present in the methods, Section 5.1, and will be addressed in Chapter 7.

1.4.2 Lack of Recorded Vegetation Features

A general problem of research efforts regarding precision agriculture in Norwegian agriculture is the lack of data on plant growth, fields, and farmer activities. While Neural Networks might be able to detect a pattern for crop growth and yield with the available features in this research study, it is far from an entirely correct representation of crop growth. This problem might present itself as a significant limitation to the study. While new features will be explored in the methods, Section 5.1, the problem will be addressed in detail in the discussion, Chapter 7.

1.4.3 Computational Requirements

The works by Engen et al. [12, 13], which is the foundation for much of this research effort, applied over 700 numeric features, alongside 30 satellite images for each farm, for a single year. For a Neural Network with many layers to process this amount of data requires a certain amount of resources and time. This can significantly build up when additional years of data or additional features are

[4] https://www.sentinel-hub.com/

applied. This research effort will be performed in a limited time frame of 5 months, with even less time available for the implementations, which will require us to prioritize the workload. Therefore, careful selection of experiments and model customization will need to happen, as the models can only be fully trained a limited number of times. This will also apply the pre-processing of new features, especially if weather interpolation is required, as this can take as much as 24 hours for one feature.

1.4.4 Identifying Noise in Satellite Image

When Engen et al. [12, 13] acquired the satellite images for KORNMO, they excluded approximately the ten most cloudy images. Still, it has been noticed that multiple images for a farm in a season tend to contain high amounts of clouds or be affected by an error in the satellite sensor, such as a wrong angle or atmospheric disturbance. Examples of these types of noise are visualized in Figure 1.1, which shows four types of noise in satellite images, all identified at the same farm in the same year. This gives some insight into noise frequency, which can be a problem when extracting information and features from satellite images and generally result in a noisy dataset that lacks essential information. This challenge will be addressed in Section 2.2.1 and discussed in Chapter 7.

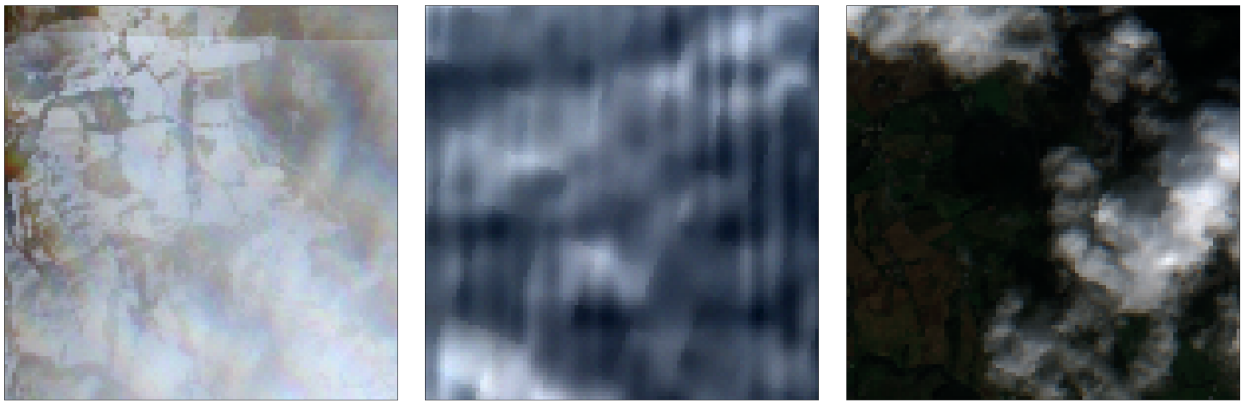

Figure 1.1: Examples of different types of noise in satellite images

1.5 Research Contributions

This research effort aims to explore and create innovative solutions for agriculture in Norway using information technology. The hypotheses presented earlier in this chapter i.e., in Section 1.3 will be researched to attain relevant information and solutions, ideally showing that the hypotheses hold true. The following contributions are proposed:

- Improving a system for grain yield predictions and monitoring for Norwegian farms, and showing that this system improves as more seasons pass.
- Proving the importance of soil quality and its relation to grain yields.
- Developing a system for grain type classification on Norwegian fields.
- Developing an accurate and realistic system for performing grain yield predictions on Norwegian farms during the growing season.
- Identifying and addressing the current state and limitations on precision agriculture in Norway.

1.6 Thesis Structure

Chapter 1 introduced the state and importance of food production and agriculture and how this field can be improved through AI. A short background on the KORNMO project was presented, along with the motivations for this research effort. Based on this, a set of hypotheses were defined with the primary goal of strengthening the position of Norway's agriculture in terms of yield monitoring and predictions. Solving this presented a set of challenges which were explained, followed by the final contributions of this research effort.

Chapter 2 provides relevant supporting background information for techniques, terms, and experiments applied in this research effort. This chapter serves as information for readers who possibly are not that experienced with agriculture or AI. Firstly, theory and research efforts revolved around plant growth factors and agricultural practices are presented. Followed up by theory relates to remote sensing and vegetation indices, along with its promising use-cases. Lastly, the background with supporting theory is given to the different AI techniques and methods applied in this research effort, such as Convolutional Neural Networks and Multiple Instance Learning.

Chapter 3 presents the state-of-the-art research effort related to each of the hypotheses. This includes the latest grain yield predictions, including in-season predictions and crop classification and mapping using satellite images and vegetation indices. These research efforts will either inspire techniques and methods applied in this research effort or be used as comparisons to evaluate our results. The critical analysis throughout the chapter evaluates how the literature and findings can be applied to this thesis's methods to perform the experiments and evaluate our hypotheses.

Chapter 4 presents the types and sources of data applied in this research effort and the handling and pre-processing of this data. This includes spectral-temporal remote sensing data, geographical data such as properties and coordinates, five different temporal meteorological features, and grain production data. This chapter will explain our data sources and why and how this data was applied to our research efforts. This data is presented in a dedicated chapter, as many of our efforts revolve around pre-processing the data.

Chapter 5 explains the experiments and methodology applied in this research effort to validate the research hypothesis. Experiments related to the two first hypotheses on improving yield prediction are initially performed through our baseline approach. This is followed by multiple experiments on classifying grain types on fields using remote sensing information and is related to the third and fourth hypotheses. Finally, all previous research efforts are combined to perform in-season yield prediction by classifying fields and processing features prior to the prediction.

Chapter 6 presents the results acquired through the research efforts and experiments performed in Chapter 5. The chapter is structured in the same way as Chapter 5. The presented results are interpreted and compared to previous related results before a detailed discussion and comparison in the next chapter.

Chapter 7 contains our interpretations, discussions, reflections and further thoughts mainly related to the results achieved in this research effort. The discussion also assesses challenges and limitations that occurred, and especially those as a result of Norway's agriculture and its lack of recorded data. Lastly, a set of suggested future works are proposed.

Chapter 8 wraps up this research effort by concluding the work and presenting our research contributions. Adding more features and data samples to yield prediction models improves the model while having its limitations. Further, grain classification is presented as being feasible but difficult with the general lack of labels. Finally, in-season yield predictions are concluded to be accurate to a certain level while only using available data at the time of the prediction. Therefore, yield prediction can be implemented into Norway agriculture as a real-time prediction and monitoring system.

Chapter 2

Literature Review

This chapter provides supporting background information on theory and techniques related to our experiments while also providing an overview of previous experiments and research efforts related to AI in agriculture. A wide range of research efforts and experiments will be explored in this chapter. Firstly, theory and research revolving around agriculture, growth factors, and farming activities will be presented in Section 2.1. Remote sensing is an essential technique throughout this research effort, as it will be applied to all five research hypotheses. A set of previous research efforts and general remote sensing findings will be presented in Section 2.2, along with methods for cloud removal. Lastly, the supporting theory will be given to the main AI techniques applied in this research effort, along with general work within yield prediction and crop classification.

Throughout the chapter, at the end of each main section, a table is attached, showing the relevant research papers explored in that section, their objectives, research design, and findings. These papers were researched in order to achieve an understanding of previous techniques and methods, along with their findings. This knowledge supports choices made and improves our methods while pushing the research efforts towards innovation.

2.1 Growth Factors in Norwegian Agriculture

Precision Agriculture is a farming system designed to increase efficiency, productivity, and profitability, much similar to the goals of KORNMO. It is a concept of using information technology to improve and optimize production while minimizing its impact on the environment and wildlife. It is often referred to as a farm-based management strategy. A Precision Agriculture system comprehends all the information available on the farm to optimize all activities and production lines. Precision Agriculture applies to several domains within agriculture, such as grain production, husbandry, forestry, and fisheries. However, like the rest of our research effort, the focus in this section will be revolved around grain production [23].

The variability in grain yields can mainly be categorized into two types:

- **Spatial Variability** - "The variability in soil, crop, landscape and environmental attributes that occurs across a certain area" [23, p. 121]

- **Temporal Variability** - "The variability in soil, crop and environmental attributes that occurs within a certain area at different measurement times" [23, p. 121]

2.1.1 Norwegian Agriculture

Norwegian agriculture is strongly dependent on farm geographical location. Most continuous areas suitable as fields are located along the coast and in the region of Innlandet, which is known to be flat compared to other regions. Figure 2.1 represents the distribution of agricultural products produced in Norway. Grain productions are mainly prioritized in the lowlands in the south of Norway, such as Innlandet. However, most land suitable for agriculture is used for husbandry and grass production, as the geographical location and climate are more suitable for this. For grain production, as visualized in Figure 2.1, barley stands for half of the grain production while wheat and oat account for about a quarter each [24].

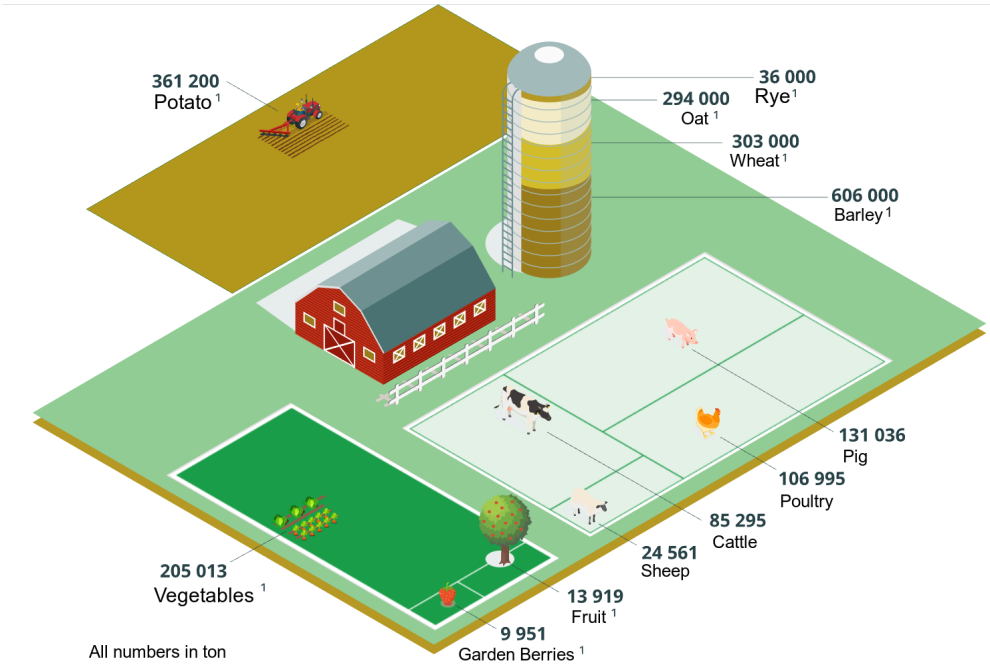

Figure 2.1: Production of agricultural products in Norway during 2020, as presented by Statistics Norway [19]

The fieldwork preparations and sowing happen between April and May in Norway, depending on how long and cold the winter and spring have been. These spring seeded crops usually reach maturity in September. They are harvested at this point, but this is highly dependent on the time of sowing and the temporal factors of the growing season. Autumn grain seeds are usually sown between August and September in Norway, but these grains are generally excluded from this research effort [25].

2.1.2 Meteorological Factors

In the spring, farmers require a period with minimal precipitation and enough temperature for the ground to thaw so that their heavy equipment can be used on the fields. Having dry soil can also make the field more manageable to work with during activities such as plowing, harrowing, and sowing. When the seeds are planted, fluorescent cool white lights are perfect for the seeds to germinate, which is when they grow into a seedling. On the other hand, incandescent lights (high in red and orange specter) are often too hot for crops and rarely suitable. After the right temperature and light, grain seeds and plants need a stable temperature and regular precipitation over the season in order to grow into harvestable grains. When harvestable, the grain plant needs to dry to a point where there is little moisture left, mainly so the harvester can process them and store grains safely [26, 25].

Temperature and precipitation are essential for farming activities, but they are also two of the most critical features for thriving plant growth [26]. Weather variables have been shown to significantly correlate with crop yield for a long time. Johnson et al. [27] presented a set of research efforts going back a long time related to this subject. While they conclude differently based on the researcher's environment and crop types, they can all be summarized that a decent combination of hot temperature and rainfall increases the yield. However, it is also vital for the crops not to have too much heat.

Sunlight can influence the development of plants in a range of ways, based on the light's intensity (amount of energy), composition (quality of the light), and duration [28]. Sunlight has also been essential for crops and specifically crucial for grains. The fields with the highest amount of sunlight are often used for growing grains, and areas with more prolonged sunlight are more likely to tackle lower temperatures [25]. While a lot of sunlight energy is good for plants, and temperature might implicitly indicate how much sunlight the plants have received, plants do not want too high a temperature. Temperatures over 30°C are stressful for most plants, and when temperatures start to reach 45°C, it becomes lethal for most plants. Therefore, a steady amount of a suitable temperature with as much sunlight as possible is preferred to maximize growth [28].

Kukal & Irmak [29] performed an extensive research effort on the effects of climate changes on maize, sorghum, and soybean yields in the U.S. The researchers found that climate variability, and more specifically, variability in precipitation and temperature, stood for a quarter of the variability in crop yields. An additional finding was that precipitation was beneficial for predicting the yield of all three crop types, while the temperature was only beneficial for some. This further strengthens crop growth and yields correlation to precipitation and temperature, showing the importance of applying these features for yield prediction.

2.1.3 Field and Soil Factors

Soil texture and quality are related to plant growth in terms of the amount of nutrients and water that can be stored, how water flows through the soil and the soil's resistance to disruptive forces. This soil texture and quality are not necessarily static. They might be negatively impacted by factors such as animals, fieldwork with machines, loss of organic matter over time, and disruptions through cultivation. These factors often affect the soil's water and oxygen capacity and lead to evaporation, severely affecting how seeds and plants thrive. Correlations for yield can also be found regarding the soil's susceptibility to water and wind erosion. Organic matter in soil helps stabilize the soil structure and bind soil particles together while acting as improved storage for nutrients and microbes. Soil water and moisture content have accounted for as much as 50% of the yield variability. Minimal water content can affect the temperature of the soil, which then affects the uptake of nutrients. All of these soil factors have been shown to relate to plant growth. Ensuring the soil has a suitable amount of water and organic matter can help plants thrive, which is why farmers might choose to fertilize or irrigate their fields [23].

Field elevation and slope changes can implicitly impact crop yields. Fields at higher altitudes are often shallower and have a coarser texture, which was shown above to harm plant growth. However, fields at a lower altitude can experience colder air, limiting the ability for water to accumulate and move through the soil [23].

Foerster et al. [16] performed crop classification and mapping and found the temporal seasonal NDVI to have a high variability at multiple time instances. This resulted from the crop's dependency on specific soil properties and the amount of water. This further strengthens the fact that crop growth is somewhat dependent on the state and properties of the soil. The researchers concluded that applying this information can improve crop classification models and possibly improve yield prediction models.

2.1.4 Farming Activities

The amount of organic matter, water, and soil nutrient storage has affected plant growth. Many farmers, therefore, tend to fertilize their fields using some organic or artificial product, adding organic matter and nutrients to the soil, thereby improving the conditions for the plant to thrive. Because of these factors, farmers might also choose to irrigate their fields based on their conditions. Suppose consecutive days of high temperature and no precipitation are recorded, and this information is applied for yield prediction. In that case, the model is likely to identify this as negatively impacting the yield. While it is generally true that this phenomenon is negative, farmers might identify this and choose to irrigate the fields, drastically decreasing the heat's negative impact. It is, therefore, important in terms of machine learning that farming activities such as fertilization and irrigation are used, as these are essential techniques applied to improve crop health and increase yield [23]. In addition to the importance of temperature and precipitation, Kukal & Irmak [29] found irrigated fields to be more robust, resulting in a more stable yield while countering climate changes. This further strengthens why farmers would like to irrigate their fields and why such information should be applied for yield predictions.

Some farmers prefer to run their production organic. Organic farming is a more natural type of farming created to maintain human health, healthy soils, and sustainable ecosystems. This type of farming tends to prioritize sustainability before yield. It is the process of combining traditional techniques with science for the human and environment's common good. Examples of this can be farmers fertilizing using organic products such as feces from animals instead of manufactured products. Another example is that organic farmers tend to switch between growing corrosive and nourishing plants on a field so the soil maintains its fertility and nutrients. These approaches are usually more healthy and sustainable, but they can also lessen the final yield of crops [30]. Therefore, knowing whether a farmer performs organic farming or not is a factor indirectly related to growth.

In order to prepare the fields for sowing, farmers tend to perform a range of field activities. First, the farmer usually plows the top surface of the field well down into the soil. This pushed any plants on the surface down into the soil, giving the soil nutrition while exposing new soil to the surface. This is done to limit non-grain plants from growing alongside the grains, as this can be too much competition. Plowing is sometimes performed during autumn, after harvest. After plowing, farmers usually cultivate their fields, spreading the new surface soil around while loosening up roots from the previously planted crops. When the seeds are to be sown, it is vital that the soil has thawed and that there is no water on the surface. Seeds are usually planted two to three centimeters below the surface for the most optimal growth. After sowing, one tends to perform drumming (referred to as *tromling* in Norway), which is about closing the sowing holes in the soil by rolling a heavy object over the field. This seals away the season, preventing birds and other animals from getting to the seeds before they can spring [25, 31].

Shah et al. [32] studied what effect different sowing dates had on the yield. They found that the sowing day significantly affected grain yields by analyzing several sowing dates and varieties. Their study is difficult to compare to Norwegian agriculture as it was located in Pakistani, where they found the 1st of November to be the most optimal sowing day. However, the researchers concluded that the sowing time tended to correlate with the yield.

This section can be summarized that many different farming activities can affect how crops thrive and affect the final yield. Therefore, many of these activities should be accounted for in the dataset when predicting yield.

2.2 Remote Sensing

Before access to Remote Sensing was possible, researchers utilized surveys of weather, farmland, and crop types for crop yield prediction and monitoring. However, this kind of data is more cumbersome to collect and can be challenging to get a hold of, especially for developing countries. Precision agriculture often requires a significant amount of samples. This type of information regarding the fields, surface, and soil and how they change is often time-consuming and costly to document and is therefore limited. This is especially true for temporal data regarding how field areas change over time. These are the reasons why remote sensing is a promising and successful technique in agriculture [5].

Remote sensing is the process of detecting and monitoring the physical characteristics of an area. This monitoring is done by measuring its reflected radiation, which is usually done by sensors on satellites or airplanes taking images of large areas on Earth [6]. One concept of remote sensing is visualized in Figure 2.2, where the Sentinel satellite system is orbiting the Earth. It can rapidly, frequently, and inexpensively collect a large amount of consistent data from regions worldwide. The more advanced systems, like the ones collecting Multi-spectral satellite images, can provide more detailed information about crop growth and farmland vegetation in places worldwide [5]. The only significant downside of remote sensing systems, once implemented, are their weakness to atmospheric changes and clouds covering the target area [33, 34]. Noise is, therefore, something that should be addressed when applying remote sensing data.

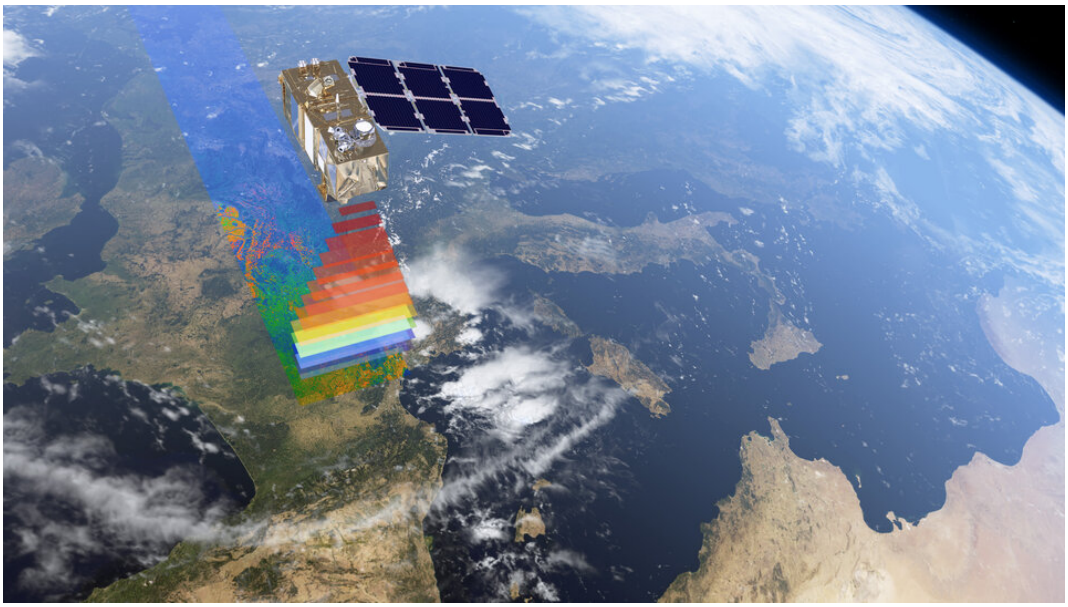

Figure 2.2: Visualization of the Copernicus Sentinel satellite orbiting the Earth while acquiring multi-spectral imagery [35]

Remote sensing technology such as satellite sensor systems can provide spectral, spatial, and temporal information related to its target objects, such as fields and farms. Spectral data is information acquired from different sensor bands and visual wavelengths, while spatial data is information concerning geographical area or location. Temporal data is information representing a state in time, meaning that the data source yields different information as time passes. Having access to information for all of these domains enables a range of different use cases. Remote sensing information

like this can be applied in agriculture to improve weed and water management, map crops and soil properties, detect drought and evapotranspiration, and identify growth status [36]. Spectral bands of remote sensing have also been used to devise as many as 519 unique vegetation indices, representing information such as chlorophyll, water, and leaf area index, among others [37].

Remote sensing has previously been used in precision agriculture, yield prediction, irrigation, weed detection, and crop and land classification. This has been done using various sensors and data collection systems, such as satellites, unmanned ground vehicles, unmanned aerial vehicles, and sensors. Because of the vast potential of remote sensing, many research papers have been published, and the interest in agriculture has exponentially grown to a point where we now have a solid level of knowledge of agricultural applications [6].

It is our belief that exploiting the information in remote sensing images covering Norwegian farmland areas can improve precision agriculture in many ways instead of relying on data manually collected and updated.

2.2.1 Cloud and Noise Removal

Data acquired by remote sensing systems are sensitive to the sensor's angle and position, along with atmospherical changes and factors such as cloud coverings. Therefore, satellite images are frequently contaminated by noise, clouds, and dead pixels [38, 14]. The reflective information is often affected or destroyed by water droplets in the air, which cause clouds, or by optical covering, which results in shadows [39]. This can severely affect the available information from remote sensing and limit the utilization of the data. Therefore, processing satellite images in an attempt to remove noise can improve any research effort performed using these images [40].

Kussul et al. [14] worked on classifying land areas using satellite images. These images contained noise and reflections from factors like clouds and shadows. To remove this, the researchers first masked out the noisy areas and applied self-organizing Kohonen maps [41]. Each map restored the selected pixels in a specific spectral band by first training the map on the whole specter and then predicting the missing values. This method could be applied to the satellite images in this research effort to improve yield predictions and grain classifications.

In addition to cloud removal in a classification-specific environment, different information reconstruction methods have been successfully performed for remote sensing images [38]. Zhang et al. [42] developed a deep learning framework that applied spatial-temporal patching to fill thin and thick cloudy areas and shadows in Sentinel and Landsat images. Lin et al. [43] chose another approach and created a technique for information cloning. Cloudy regions are replaced with regions with well temporal correlation from cloud-free remote sensing images. This was successful with the assumption that the land cover information from remote sensing images would not change significantly in-between short periods.

The methods presented above could only perform a single type of information reconstruction task. Zhang et al. [44] improved on this by developing a novel method for reconstructing missing information in remote sensing features. The researchers used a deep CNN combined with spatial-temporal–spectral information to create a framework with the ability to reconstruct dead pixels and remove clouds from remote sensing images. Their framework showed promising results in solving these tasks relevant to cloud and noise problems in this research effort. Applying this framework to the satellite images of Norwegian farms requires datasets consisting of cloudy and cloud-free sets of spatial-temporal data. Transforming the raw satellite images of Norwegian farms to a suitable training dataset will require considerable time.

2.2.2 Derived Vegetation Indices

Remote sensing systems can capture and interpret the spectral composition of reluctance emitting from the Earth's surface. This data contains information about the physical properties of land, water, and different vegetation features. Vegetation indices are simple transformations computed directly from two or more remote sensing spectral bands. Vegetation indices are designed to convert the range of reflective information into something smaller, concrete, and more interpretable to enhance vegetation properties and terrestrial variations. They usually cover a specific domain or type of information, such as the amount of water or leaf size [45, 46]. Pearson and Miller [47] created the two very first vegetation indices in 1972 to learn the contrast between vegetation and the ground surface. Since then, more than 519 unique vegetation indices have been developed for different purposes, use-cases, and environments [37]. Therefore, this section will present a set of general works using vegetation indices, followed by a list of the vegetation indices applied in this research effort and their background.

Considering the high amount of different vegetation indices that have been developed for different unique goals in different environments, one should not apply a random set of vegetation indices. For example Xiao et al. [48] and Zhong et al. [49] found that the widely used Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) alone did not always supply the most accurate information. An essential process of crop type classification is therefore to identify suitable features and techniques for our environment and determine which to use. This is especially important for grain type classifications, where the difference between grain types is less than that of other crop types in general [18].

One of the more fundamental studies on crop mapping and classification is the works of Massey et al. [17]. Decision trees were implemented to classify and map crop types for each growing season, using time-series data of the Normalized Difference Vegetation Index (NDVI) extracted from 250-m spatial resolution satellite images. The researchers achieved an accuracy as high as 78%, proving that NDVI could be applied to classify crop types. The researchers concluded that they were not able to achieve a higher accuracy because of the significant similarity between the target crop types.

The following are the ten vegetation indices selected for crop classification in our research efforts. These indices were selected based on acquired knowledge of theories, techniques, and research efforts and are presented in this section along with their theory, application, and formula. The formulas combine Kobayashi et al. [50] and Henrich et al. [37] and the references used for each index. As further knowledge, the wavelength used in the formulas is not meant to be exact, as different remote sensing systems do not have the same wavelengths for the same type of bands. This is why many of the formulas list their wavelength as a range with minimum and maximum values for each band type. Using the neighboring band wavelengths for some vegetation indices is also possible.

Enhanced Vegetation Index (EVI) distinguishes canopy structure and vegetation growth status of crops and is calculated from visible near-infrared. Equation 2.1 shows how EVI is calculated. L is a canopy background adjustment value. The C_1 and C_2 are aerosol resistance coefficients that correct aerosol influences between wavelength *(640:760nm)* by using a band between wavelength *(420:480nm)*, and G is the gain factor. These factors are used between different remote sensing systems to account for their unique reflections and atmospheric values [46].

$$G * \frac{(780:1400nm) - (640:760nm)}{((780:1400nm) + C_1 * (640:760nm) - C_2 * (420:480nm)) + L)} \tag{2.1}$$

Land Surface Water Index (LSWI) distinguishes the water content in crops based on near and shortwave infrared. These bands are sensitive to soil moisture and leaf water, making LSWI sensitive to water. Equation 2.2 shows how LSWI is calculated [48].

$$\frac{(841:875nm)-(1628:1652nm)}{(841:875nm)+(1628:1652nm)} \tag{2.2}$$

Normalized Difference Senescent Vegetation Index (NDSVI), designed by Qi et al. [51], distinguishes both water content and the growth status and is based on visible shortwave infrared, a combination of EVI and LSWI. Equation 2.3 shows how NDSVI is calculated. These wavelengths are in the water absorption regions of the spectrum and can therefore be combined to extract information related to senescent vegetation.

$$\frac{1640nm-(640:760nm)}{1640nm+(640:760nm)} \tag{2.3}$$

Normalized Difference Vegetation Index (NDVI), designed by Rouse Jr et al. [52], measures the amount and density of green vegetation based on near-infra-red and red bands. This is useful for Agriculture, as unhealthy crops reflect less near infra-red than healthy crops [53]. Equation 2.4 shows how NDVI is calculated. The normalization in NDVI enables it to measure the greenness more consistently with fewer deviations than a more straightforward ratio of the two bands [52].

$$\frac{(1300:3000nm)-(800:900nm)}{(1300:3000nm)+(800:900nm)} \tag{2.4}$$

Shortwave Infrared Water Stress Index (SIWSI) estimates the water content of crop leaves. A short wave infra-red band operates on a wavelength influenced by leaf water content, enabling SIWSI to extract information related to the level of water content in crops. This can be used to identify the level of stress inflicted by the water content. Equation 2.5 shows how SIWSI is calculated. An index value below zero indicates that the crops have sufficient water content. In contrast, values above zero indicate that the crops are experiencing some level of water stress due to too much water [54].

$$\frac{860nm-1640nm}{860nm+1640nm} \tag{2.5}$$

Green-Red Normalized Difference Vegetation Index (GRNDVI), created by Wang et al. [55], is one of multiple vegetation indices using a combination of red, green, and blue bands in an NDVI manner. GRNDVI is one of the better vegetation indices for measuring Leaf Area Index (LAI), an important feature related to crops health, growing stage, and type. Equation 2.6 shows the calculation for GRNDVI.

$$\frac{(780:1400nm)-((490:570nm)+(640:760nm))}{(780:1400nm)+(490:570nm)+(640:760nm)} \tag{2.6}$$

Normalized difference Red-Edge (NDRE) is a vegetation index similar to NDVI, which uses a ratio between near-infra-red and red-edge bands. It can extract information from remote sensing data regarding the health and status of crops, including the greenness of the canopy. Equation 2.7 shows the calculations for NDRE [56, 57].

$$\frac{(780:1400nm)-(690:730nm)}{(780:1400nm)+(690:730nm)} \tag{2.7}$$

Structure Intensive Pigment Index 3 (SIPI3) is a function of carotenoid and chlorophyll a. Carotenoid provides information related to canopy physiological status while chlorophyll provides information related to plant photosynthesis. The original SIPI was designed by Penuelas et al. [58] while attempting to achieve the highest correlation between carotenoid and chlorophyll for acquiring plant physiology and phenology information. SIPI3 is a variation of this, shown in Equation 2.8.

$$\frac{800nm-470nm}{800nm-680nm} \tag{2.8}$$

Photosynthetic Vigor Ratio (PVR) is a simple combination of the red and green bands, reflecting chlorophyll absorption, canopy greenness, and photosynthetic activity. Equation 2.9 shows the formula for PVR [57, 23].

$$\frac{550nm-650nm}{550nm+650nm} \tag{2.9}$$

Green Atmospherically Resistant vegetation Index (GARI), created by Gitelson et al. [59], is a vegetation index resistant to atmospheric effects while also being sensitive to a range of chlorophyll-a concentrations, plant stress, and rate of photosynthesis. GARI is as much as four times less sensitive to atmospheric effect than NDVI while also providing much of the similar information. Equation 2.10 shows how GARI is calculated.

$$\frac{(780:1400nm)-((490:570nm)-((420:480nm)-(640:760nm)))}{(780:1400nm)-((490:570nm)+((420:480nm)-(640:760nm)))} \tag{2.10}$$

2.2.3 Overview of Remote Sensing Research Efforts

Table 2.1 below presents a list of the research efforts on remote sensing with contributions highly relevant to our work. These are shown with their objectives, research design and findings.

Authors	Objectives	Research Design	Findings
Zhang et al. [42] (2020)	Determined and solved a number of problems related to cloud-removal in remote sensing images	A spatial-temporal patching approach to cloud and shadow removal in temporal satellite images	Developed a framework suitable to remove or replace thin and thick clouds, and shadows
Lin et al. [43]	Removed cloud-contaminated regions from temporal remote sensing images	A cloning approach to cloud removal where contaminated regions are replaced by clean regions with high correlation	Developed a technique able to process large clouds in a range of different landscapes
Zhang et al. [44] (2018)	Improve cloud and noise removal techniques to perform multiple reconstruction tasks at once	Deep CNN applied on spatial-temporal-spectral for multi-task reconstruction	Created a framework for correcting dead pixels, scan line corrector-off problem and thick clouds
Massey et al. [17]	Crop classification in U.S using NDVI	NDVI time-series data derived from MODIS satellite system, applied to decision trees	Developed a generalized and a year-specific crop classification decision tree which achieved 75% and 78% accuracy respectively

Xiao et al. [48]	Paddy rice mapping in Asia by using remote sensing data	Calculated time-series of vegetation indices from MODIS satellite images, for identifying flooding and transplanting in rice fields	Developed a geo-spatial dataset for paddy rice fields in Asia while discovering that NDVI and EVI not always contributes to crop classification
Zhong et al. [49]	Classify and map corn and soybean in the U.S.	Calculated phenological metrics from Landsat-derived EVI time series by using image segmentation and curve-fitting	Developed suitable phenological metrics for crop classification which achieved 88% accuracy on Random Forest. Also discovering that EVI not always contributes to crop classification
Huete et al. [46]	Analyzed the performance of NDVI and EVI in terms of radiometric and biophysical perspectives.	Derived NDVI and EVI from satellite images of a whole year and analyzed the correlation between the indices vegetation differences. This was also evaluated for multi-temporal indices.	Identified a correlation between NDVI & EVI, and top-of-canopy reflections. Found NDVI to saturate in regions with high biomass, while the EVI were still able to identify canopy variations.

Table 2.1: Overview of explored research efforts related to general remote sensing

An artificial Neural Network (ANN) is a computing model consisting of several neurons interconnected someway with each other and is based on the concept of the perceptron. A perceptron is an information storage model similar to the brain, with neurons, like a nervous system. The system works by having each neuron perform some simple transformation or mapping of its inputs and forward the output to some other neuron. These Neural Networks usually consist of sub-components called layers, which can be represented as a perceptron, containing a variable number of nodes and is represented as a function. These layers perform some specified computation on their inputs and then forward the outputs to the next layer when activated through an activation function. Each node has an adjustable variable called weight, which impacts the computation and turns the network into a variable, learnable network. An ANN takes a set of inputs and runs them through each layer up until the last layer, where they are transformed and outputted. The layers in-between the network's input and output are typically referred to as hidden layers. While each node is simple, the whole network can become highly complex when enough are connected. A common approach, which is also used in this project, is to extend this kind of network with multiple hidden layers, turning it into a Deep Neural Network (DNN) [60, 61].

Typically, the steps performed to train a neural network after defining its architecture are described as the following [62]:

- Initialize the neural network with adjustable parameters, usually to a random state.
- Apply a set of input features to the network.
- Process the inputs by computing them through each layer, where the last layer produces the output labels.
- Compute the loss based on how far off the output labels are from being correct.
- Calculate the gradient loss concerning all weight in the network through backward propagation.
- Adjust the weights in the network based on the learning rate and the calculated gradient loss. This adjustment and correction are what "trains" the network.

These steps adjust the weights on each node in the network and are supposed to enable the model to perform more accurate calculations than the previous iteration.

2.2.4 Adam

Adam is an first-order gradient-based algorithm for optimization of stochastic functions. It is a efficient and commonly used optimization algorithm, especially for problems related to big datasets, such as out satellite images [61, 63].

2.2.5 Softmax

The softmax function takes a vector as input and normalizes it into an exponential probability distribution. This is often used as the activation function on the last layer of neural networks to output a probability distribution in multi-class machine learning problems [61].

2.2.6 Dense Layer

A Dense layer is a layer commonly used in ANNs, that is deeply connected with its following layer. Deeply connected means that every neuron of the layer is connected to every neuron of its preceding layer. The operation performed in a dense layer is a matrix multiplication of inputs with weights. Additionally, one can supply the layer with an activation function that performs this function element-wise. Given the rules of matrix multiplication, the output of a dense layer is a vector of size x. Therefore, the dense layer can be used for changing the dimensions of vectors [64].

2.2.7 Dropout Layer

A Dropout Layer is a layer that takes the entire output vector from the previous layer as input and randomly sets input units within the vector to 0. The rate at which units are changed to 0 is an adjustable input parameter between 0 and 1. Additionally, all other inputs are scaled up by a rate of $1/(1 - rate)$ such that the sum over all units in the vector is unchanged. Dropout is a typical method for preventing overfitting [65].

2.2.8 Flatten Layer

A Flatten layer is simply a layer that takes a vector as input and reshapes the vector to a shape in the form $(None, x)$. For instance, if a vector shaped (None, 10, 5, 64) is inputted to the flatten layer, the output vector will be (None, 3200) [66].

2.2.9 Batch Normalization Layer

Batch Normalization is a transformation that maintains the mean output close to zero and the output standard deviation close to one. This process functions differently during training and inference. Inference refers to the process in which the model is evaluated or is predicting data. During training, the layer normalizes its output using the mean and standard deviation of the current batch of inputs. During inference, the layer normalizes the output using a moving average of the mean and standard deviation of the batches it has seen during training [67].

2.2.10 Mean Absolute Error

Mean Absolute Error (MAE) is an evaluation metric used for regression models. The MAE is calculated by taking the mean of all absolute errors. Each absolute error is calculated as the absolute value between predicted and true values. The following formula shows the calculation of MAE:

$$mae = (\frac{1}{n}) \sum_{i=1}^{n} |y_i - x_i|$$

where n is the number of samples, y_i is the true value of sample number i, and x_i is the predicted value for sample number i.

2.2.11 Overfitting

Overfitting is a phenomenon occurring when a machine learning model performs significantly better on its training data compared to new and unknown data. Therefore, a model that only learns patterns within the training data is unable to learn anything from new data. Overfitting can be minimized by augmenting the training data, reducing the model size, including dropout layers in the model, and implementing weight regularization. [61]

2.2.12 Hyperparameters

Hyperparameters are tunable variables that one may use to customize existing machine learning libraries to their favor. The hyperparameters attempt to account for every unique combination of machine learning features and environments, controlling the outcome of an algorithm or learning process. The following is a list of frequently used tunable parameters relevant to this research effort:

- **Learning Rate** in machine learning is the step size in the loss function when updating and adjusting the weights. In a way, it is often the speed at which a model learns. A low learning rate in the given setting might result in the model learning slowly or even getting stuck without progress. In contrast, a too high learning rate may cause the model to overfit and learn the complete training data distribution, performing poorly on new data. Therefore, this hyperparameter is often the most important to tune correctly, as the rate and quality of the model's learning are significantly impacted by this [61, 68].
- **Hidden Layer Size** determines how big each layer in the network will be. The size number specifies how many neurons there will be for that layer. Therefore, this size determines the number of weight variables and the computation required when iterating through the network. Typical sizes for the hidden layer range from *16* to *512* but are generally in the 2^n range [61, 68].
- **Weight Decay** is used as a penalty to the loss function to limit large weight being stored in the model and overfitting. If the weights grow too large, minimizing the weight might be prioritized more than minimizing the training loss [61].

2.2.13 Convolutional Neural Network

A Convolutional Neural Network (CNN) is a Neural Network (NN) designed to process multi-dimensional data. Its architecture is optimized to reduce the computing time when learning weights and biases for this data. CNNs are split into three layers:

1. Convolution layers
2. Activation layer
3. Pooling (Optional)

The convolution layer consists of several filters, each with its own trainable weights of size m x n. These filters are applied to the entire input in a sliding window fashion, as shown in Figure 2.3. A typical filter size is 3 x 3, meaning that each pixel in the output from the layer is created by looking at a grid of 3 x 3 pixels directly above it. The layer also contains a stride which determines how fast the sliding happens. With a stride of 1, the filter moves one pixel at a time [69].

Pooling is an operation of reducing data resolution, often used after a convolution layer. This is beneficial for decreasing computational requirements since it reduces the number of parameters to optimize. Pooling also prevents overfitting by generalizing features. Similar to convolution, pooling also uses a stride and a filter. The most common pooling, max pooling, is done by applying a max filter on each region it passes over, only returning the highest value. [61]

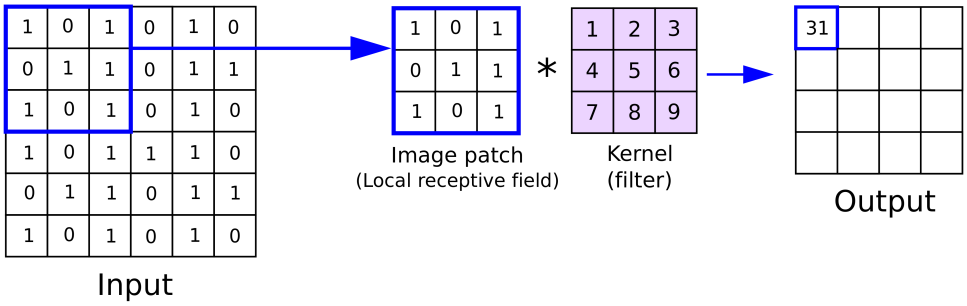

Figure 2.3: Example of convolution in CNNs [70]

Convolutional layers are often used for images because they can extract image features effectively. One property that enables this is called parameter sharing, which means that each weight is used multiple times during a single pass through the network. Another property is something called sparse connectivity. Sparse connectivity means that each layer does not have as many connections to the next layer. Each area the filter overlaps connects to one node, as shown in figure 2.3. Lastly, convolutional layers are equivariant to translation, which means that if the input is translated, the output is equally translated. In practice, this means that the layer will detect features regardless of their position in the data. [61]

The layers in a convolutional neural network are often laid out to become smaller and smaller in size but larger in depth as the data traverse deeper into the network. The first layers often extract low-level features like edges and color tones, while the later layers often extract high-level features like shapes and objects. This property makes them able to extract most of the features present in an image [61]

2.2.14 Recurrent Neural Network

A Recurrent Neural Network (RNN) is a class of neural network that uses sequential data as input instead of the traditional flat data structures and matrix data used in standard neural networks. The unique property of RNN is its ability to remember weights and patterns from previous timesteps. This is achieved by passing the output of the previous timestep alongside the input at the current timestep. Each timestep is processed using the same weights as all other timesteps, allowing RNNs to generalize well on sequences of varying lengths [10].

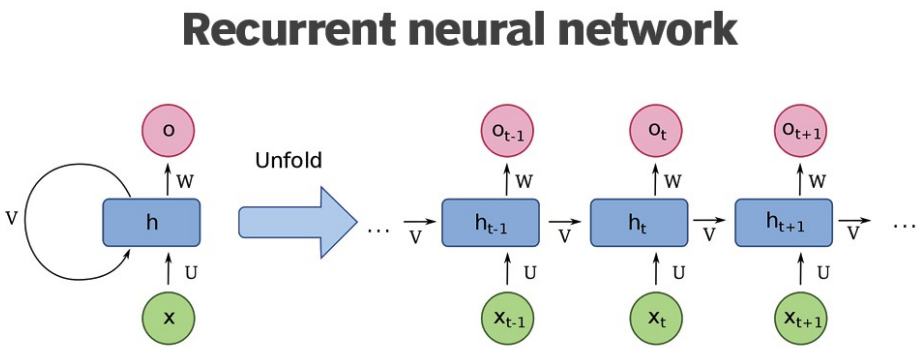

Figure 2.4: Example of Recurrent Neural Netowrk [71]

As RNNs pass on the output through the model, it feeds a hidden state (v) back to itself, as shown in figure 2.4. Each step also produces an output (o) and a hidden state (v). The hidden state is continually changed by new information at each step and can be viewed as an encoded representation of all previous steps. Two common types of RNNs are Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). Given that we only used GRU during this research effort, the forthcoming section only describes GRU.

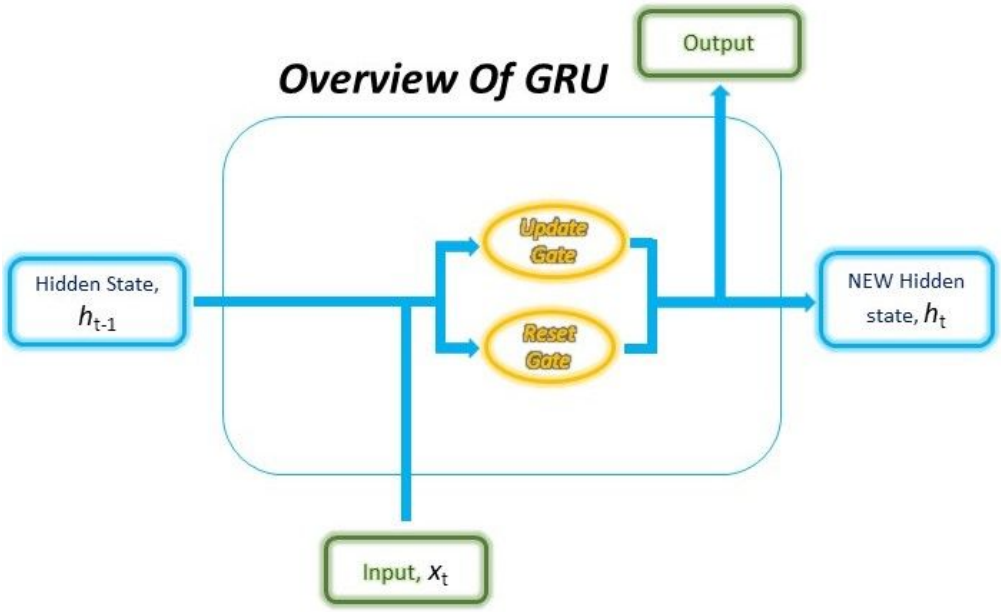

Figure 2.5: Example of a Gated Recurrent Unit [72]

The Gated Recurrent Unit is a new variant of LSTM and was designed to have simple computations and implementation. Compared to a regular RNN partition, the GRU has two gates: a reset gate and an update gate, as shown in figure 2.5. The reset gate determines which parts of the previous state should be ignored, and the update gate decides which parts of the hidden state should be updated. These mechanics allow the GRU to drop information that is no longer relevant, which

helps RNNs remember long-term information [73].

2.2.15 Neural Networks for Agricultural Purposes

Yield Prediction using Neural Networks

Crop yield prediction has previously been performed in various environments, using different data types, and applied to a range of machine learning models, as shown in this section.

One of the more straightforward research efforts on yield prediction using remote sensing is the works by Raun et al. [74], who performed a research effort to determine if grain yields of winter wheat could be predicted using multi-spectral seasonal features. The researchers found a high correlation with yield by computing two NDVI measurements into an estimated yield feature. This feature explained 83% of the yield variability but could not account for growth changes due to rain, lodging, and shattering. Nevavuor et al. [75] and You et al. [5] attempted to apply deep learning for crop yield predictions. The former applied RGB bands and an NDVI band from multi-spectral images of fields to a CNN model for wheat and barley yield predictions. Their model reached a Mean Absolute Percentage Error (MAPE) as low as 8.8%. The latter applied temporal satellite images as histograms to CNNs and LSTMs along with textual features such as soil type and fertilizer rate to predict soybean yields, reducing root-mean-square-error (RMSE) by 30% from previous U.S surveys.

Khaki & Wang [76] applied genotype, soil, and weather data to a DNN to predict corn yield and yield difference. The soil data consisted of eight features covering information related to clay, silt and sand percentages, water, and pH level, while the weather data consisted of 72 features regarding day length, precipitation, solar radiation, vapor pressure, and maximum and minimum temperature. Training the DNN on these types of features resulted in a RMSE of 11%. This outperformed previous research efforts on Regression Tree [77] and Least Absolute Shrinkage and Selection Operator [78], indicating that DNN might be a superior technique for our research effort as well. Their research efforts also suggest that soil and additional weather data can contribute positively to yield prediction. They further stated that weather prediction is an essential and inevitable part of crop yield prediction. The researchers therefore also successfully predicted weather features for the current growing season using a simple Neural Network for each weather feature. They found year, day, location, and weather values from the four last years to be the most optimal features. Replacing the real weather value with these predicted values in their research effort on corn yield prediction only increased the RMSE by 1%. This also applies to our research effort, where weather features are crucial and should be predicted for early grain yield predictions.

Russello & Shan [11] applied satellite images as input to a CNN network to predict soybean yields in the U.S. Assuming that pixel position for satellite images is not important, they transformed every image over a whole year into a 3D histogram for each location. This approach determined that features between February and September were the most important for predicting soybean yields, which aligned well with the growing season of their location.

Basnyat et al. [79] researched the optimal time for using remote sensing to predict crop yield by studying the correlation between the NDVI index from three different periods and the yield of different crop types. They concluded that NDVI derived from the period between the 10th and 30th of July appeared to be the most suitable for yield predictions of spring-seeded grain that typically mature in August. More specifically, they found 26 of July to be the date where they generally experienced the highest correlation for the features. This corresponds with the theory of the NDVI index, as explained in Section 2.2.2. While keeping in mind that spring-seeded grains in Norway typically reaches maturity in September, as presented in Section 2.1.1, these findings are important for evaluating when early predictions are suitable to be performed for Norwegian farms.

Crop Classification and Mapping using Neural Networks

Satellite images have previously been used successfully to classify different crop types and have the potential to represent their spectral dynamics, which can be used to detect phenological differences between them. However, using satellite images for this type of classification has its problems [18]. For instance, Hu et al. [80] and Qui et al. [81] found that stacking satellite images over the growing season often resulted in redundant information. It is therefore important to evaluate what types of features to apply and what type of AI architectures and techniques to use.

Castillejo-González et al. [34] researched how crop classification performance changes when going from pixel-based classification (1D), to object-based classification (pixels grouped up together). They found every research effort to outperformed the pixel-based approach, indicating that 2D images should be prioritized above single pixel values for grain classification.

Ji et al. [15] applied multi-spectral temporal satellite images to 3D CNNs for crop classification. Using a 3D CNN framework, they learned the temporal representation of the whole growing season. Their model outperformed similar 2D methods and was exceptionally good at identifying growth characteristics and crop types. 3D CNN architectures have been extensively shown in this section to be superior to 2D CNN, considering feature extraction and representation of multi-spectral temporal data.

Because of the lower accuracy of pixel-based classification methods, Peña-Barragán et al. [82] developed an Object-based Crop Identification and Mapping model, which achieved an accuracy of 79%. They found that the remote sensing images from the late-summer period contributed to 60% of the classification model learning. Secondly was the mid-spring images of around 30%, and lastly, the early-summer images stood for the last 10%. This interesting difference shows how vital the late growing season images are. The findings also indicated how difficult it might be for a classification model to classify early in the growing season.

2.2.16 Attention-Based Multiple Instance Learning

In image classification research efforts, the dataset is typically labeled consistently to its respective class(es), but this is often not the case for real-life applications. A typical example of this is machine learning work on medical imaging, where multiple images might exist with only a general statement of their belonging label. This problem also exists for the crop labels for each field in this research effort. These labels, containing information regarding what is currently planted on the fields, only partially cover the available fields and farms, resulting in a semi-labeled dataset. A solution to this problem has been the extensively researched Multiple Instance Learning (MIL), a semi-supervised learning method [83, 84, 85].

MIL revolves around predicting bags' labels where bags contain a mix of labeled and unlabeled instances. This replaces the more traditional technique of classifying a single instance at a time. A single target class is used in a traditional MIL setting, and bags are labeled as either positive or negative based on whether they contain at least one target class instance. Therefore there is no need to know the label of every instance [86, 84]. This means that the MIL model uses a specific number of bags as training data and predicts whether a bag is positive or negative. Further, the attention mechanism that can be used in a MIL model is what gives each instance in a bag an attention score. The attention score is a percentage of how much influence the instance had in classifying a bag. These percentages are calculated using the softmax function, which means that the all attention scores sum up to one.

Multiple research efforts around bag classification have previously been performed. Cheplygina et al. [87] supplied a vector of each bag as a feature to the training dataset, where one vector represented that bag's dissimilarity to the other bags. Chen et al. [88] used the similarities between

instances in one bag to create a feature space, into which the whole bag was then mapped. Only the crucial features from the feature spaces were extracted and used in a traditional supervised learning approach. Ramon & De Raedt [89] and Raykar et al. [90] researched different types of instance-level classification and feature selection within bags, reaching promising and competing results.

2.2.17 Overview of AI in Agriculture Research Efforts

Table 2.2 below presents a list of research efforts on machine learning applied to agriculture. This mainly consists of general work on yield prediction and crop classification and mapping, but also work on semi-labeled dataset. These are shown with their objectives, research design and findings.

Authors	Objectives	Research Design	Findings
Raun et al. [74]	Determine if grain yield can be predicted using in-season spectral-temporal information	Calculated two types of NDVI which were used to create a yield estimate. The yield estimate were compared with the actual yield	Found an overall 83% correlation to actual yield and found NDVI to work well on sites where rain, lodging and shattering did not significantly impact the yield
Nevavuor et al. [75]	Crop yield prediction based on RGB and NDVI data	NDVI and RGB data from UAVs' are applied to a CNN with Adadelta and L^2 regularization with early stopping	Reached an MAE of 4840 kg/daa and an MAPE of 8.8% for crop yield prediction
You et al. [5]	Improve state-of-the-art crop yield techniques in terms of scalability, accuracy and cost	Based on modern representation learning ideas, they used dimensionality reduction to apply remote sensing and soil data to CNNs and LSTMs with a Gaussian Process component	Reduced the RMSE by 30% from previous techniques
Kukal & Irmak [29]	Evaluated climate and irrigation impacts on sorghum, soybean, and maize yields in different U.S counties.	Analyzed correlations between climate changes and variability in crop yields from extensive datasets from 1968 to 2013.	They found variability in temperature and precipitation to account for one quarter of the variability in yields, and found irrigated fields to be more robust against climate changes. Concluded that precipitation had a positive impact for all crops.
Khaki and Wang [76]	Predicting corn yields and analyzing the yield difference between corn sub-classes	Crop genotype, soil information and weather data applied to Deep Neural Networks for yield prediction and analysis of the importance of each feature	Achieved a yield prediction accuracy with 12% RMSE while seeing the importance of solar radiation, temperature and precipitation
Russello & Shan [11]	Predicting soybean yields in the U.S.	3D-histogram of temporal satellite images applied to CNN	Developed a 3D CNN for spatial-temporal features which outperformed previous CNN and LSTM methods
Basnyat et al. [79]	Analyzed data from remote sensing to determine the optimal time to acquire remote sensing data that related to grain yields	Derived NDVI for pea, canola and spring wheat for three different dates in the growing season and evaluated each of their correlation to the final yield.	Found NDVI to have the highest correlation to spring-seeded crops in the period between the 10th and 30th of July. In addition, they found well correlations for pea yield, but limited correlations for canola.
Hu et al. [80]	Compare two approaches of using the pairwise separability vegetation (SI) index to classify crop types	Calculated SI minimum and SI average of spectral-temporal remote sensing data and evaluated their performance for crop classification	Discovered that SI minimum generally performed better, and that using images of high temporal resolution does not necessarily increase accuracy

Qui et al. [81]	Solve the problem of cloud cover and coarse temporal resolutions on satellite images for paddy rice mapping	Paddy rice mapping in China by dividing regions per-pixel, extracting cloud-free remote sensing images, calculating phenological metrics and applying them to Random Forest	Developed an Adaptive paddy Rice Mapping Method with an overall accuracy of 95.77%
Castillejo-González et al. [34]	Analyzing the accuracy of different crop identification and classification methods	Applied multi-spectral images to five different classification methods to understand the possibility of crop classification in the Mediterranean area	Achieved an accuracy as high as 85% and found that object-based classification is superior to pixel-based
Ji et al. [15]	Crop classification using spatio-temporal data from remote sensing	Designed a 3D CNN fitting the spatio-temporal data, with active learning	Found their 3D CNN to outperform similar 2D experiments while also increasing accuracy through active learning
Peña-Barragán et al. [82]	Object-based image analysis for crop identification and mapping	Applied several vegetation indices and textual features from satellite images to classify 13 different crop types while analyzing each feature and growing months importance	Achieved an accuracy as high as 79% while finding late-summer and mid-spring to contribute to 60% and 30% of the classification, respectively
Cheplygina et al. [87]	Improve MIL by representing each bag as a vector of its dissimilarities to all other bags	Researched different ways to define dissimilarities between bags and when each of them are suitable to use	Found different ways to calculate dissimilarities, like treating bags as an attributed graph, point set or distribution based on instances. These calculations also shown to be inexpensive
Chen et al. [88]	MIL without relating a single true instance label to the bags label	Applied instance similarity measure to map bags into a space based on their instances, and then selected the important features through 1-norm SVM	Designed Multiple-Instance Learning via Embedded instance Selection with competitive accuracy and efficiency
Raykar et al. [90]	Bayesian Multiple Instance Learning for relevant feature selection	Bayesian automatic relevance determination paradigm is used to select the most relevant features in a bag and inductive transfer is used between similar classifiers	Created a more accurate MIL technique which narrowed down the total feature dataset for a bad into a small subset of useful features

Table 2.2: Overview of explored research efforts related to machine learning techniques applied in agriculture

Chapter 3

State-of-the-art

This chapter presents the state-of-the-art research efforts related to each of our hypotheses. We introduced the problem statement, motivation for our work, and the proposed research hypothesis in Chapter 1. This was followed by Chapter 2, where relevant background theories and research efforts were presented for agriculture, remote sensing, and AI. This chapter will present the most promising technique for cloud and noise removal and multiple research studies on crop classification using vegetation indices to enable the selection of indices. In addition, it will present comparative state-of-the-art techniques and solutions for early yield prediction and crop classification. The research efforts in this chapter will inspire the techniques and methods applied in this research effort, which will be used as comparisons to evaluate our results when suitable. Throughout the chapter, a critical analysis will be performed to evaluate how the state-of-the-art can be applied to this thesis's methods to perform the experiments and evaluate our hypotheses.

3.1 Removing Cloud and Noise in Remote Sensing

Detecting clouds in remote sensing images to create a training dataset might not be a viable option for this research effort. As a solution to this requirement, Wen et al. [91] developed a two-step Robust Principal Component Analysis (RPCA) framework, which detects regions with clouds in satellite images, removes the clouds, and restores the region. Their innovation resulted in a framework that does not require any cloud-free satellite image dataset or any cloud detection pre-processing. While multiple promising cloud removal method have been presented, the latest work by Wen et al. [91] presents itself as the most suitable technique. While considering the limited amount of time available, their methods could be applied as an effort to remove noise from satellite images and thereby likely improve the accuracy of relevant deep learning models.

3.2 Predicting Grain Yields During the Growing Season

We explained the importance of crop yield prediction and how it can improve sustainability and reduce agriculture costs in Chapter 1. Chapter 2 showed different research efforts on crop yield predictions related to our study. This will be extended by this section's state-of-the-art research efforts and techniques, mainly around predicting the yield during the growing season. These efforts will be our main inspiration for the early yield prediction methods and the comparisons for our results.

Sharma et al. [10] researched the importance of each month in the growing season, and found the earlier months to be the most importance once. They evaluated how satellite images from each

month of the growing season correlated with yield prediction. By replacing all images in a month with noise, they could evaluate the importance of the features from that month based on the root mean square error (RMSE). Their significant finding was the importance of the earlier months, especially the first month. Removing the satellite images from the first month increased the RMSE by as much as 700. This could result from the farming activities performed at the beginning of the growing season, further showing their importance, as explained in Section 2.1.4.

Sharma et al. [10] developed a technique for predicting wheat crop yields using raw satellite images. They applied a CNN for each satellite image in a time series to extract relevant information from the 12 bands. The output from the CNNs' was concatenated into an LSTM, which encoded the temporal features, predicting the yield. Their model achieved a training and validation loss of 0.1% and 0.15% respectively. They further improved the accuracy by incorporating additional features such as location and area information to the input vector of LSTM. Their research effort and system served as a significant inspiration for the previous work by Engen et al. [12, 13], the latest crop yield research effort for the KORNMO project. The technique used by Sharma et al. [10] will thereby also have a considerable significance in the codebase of our research effort.

As mentioned earlier in Chapter 1, a group of students has already researched grain yield prediction in Norwegian agriculture on behalf of the KORNMO project. Engen et al. [12, 13] applied multi-temporal satellite images, weather data time series, and farm information to various models. Their best performing model was a Hybrid Deep Neural Network consisting of sets of CNNs, just as Sharma et al. [10] applied, and a Gated Recurrent Neural Network for concatenating the CNN outputs with a vast amount of weather, farm, and field features. Their research efforts concluded with the hybrid model reaching an mean absolute error (MAE) as low as 76 kg/daa. With limited prior work on farm-scale yield predictions available, they compared their results with that of the study by Sharma et al. [10] and found their model to perform on a comparable level.

During the research efforts, Engen et al. [12, 13] attempted to use their best performing model to predict the yield earlier in the growing season. The researchers predicted using data from week 10 to 26 and from week 10 to 21 and found an error rate increase of 7.66% and 20.89%, respectively. This type of study is an essential prerequisite before being able to implement in-season yield prediction in a natural environment like Norwegian agriculture. Still, to the best of our knowledge, the researchers used features not available at that time in the growing seasons, such as grain types, growing area, and production benefits. This indicates the need for crop classification on Norwegian farms, as this missing information currently limits in-season yield predictions.

Engen et al. [12, 13] made use of many of the available data sources in Norway that are relevant for agriculture. This included Sentinel-2 satellite images[1], grain production and farmer information from the Norwegian Agriculture Agency[2], meteorological data from The Norwegian Meteorological Institute[3], cadastral layer from Kartverket[4], and field boundaries from Geonorge[5]. With the limited data available in Norway related to grain production, it is natural to think that many of these data sources, if not all of them, will be useful for our research effort as well. The work of Engen et al. [12, 13] on retrieving and processing data from these sources into features applicable for yield prediction will also be highly relevant and helpful. Because of this, their codebase and acquired datasets were handed over at the beginning of this research effort in order to expand on their area while also saving valuable development and computation time.

[1] https://sentinel.esa.int/web/sentinel/home
[2] https://www.landbruksdirektoratet.no/nb
[3] https://frost.met.no/index.html
[4] https://www.kartverket.no/en
[5] https://www.geonorge.no/

3.3 Classifying and Mapping Crop Types

In order to apply remote sensing for crop productivity and yield prediction, an almost mandatory task should be done in advance, which is crop identification and farm-land area calculation. As mentioned in Chapter 1, this information is currently unavailable during the growing season in Norway. The information is not even accurately available for every farmer even after the growing season. Therefore, farm-land area and crop content will need to be predicted, and only then can the actual prediction of crop yield happen [92].

Kussul et al. [14] researched how field classification performance changes when a model is allowed to learn the spatial context of an image (2D) compared to learning spectral domain from single pixels (1D). Each of the two implementations had five sets of CNNs. Each set consisted of two convolutional and max-pooling layers, followed by two fully connected layers with a varying number of neurons for each CNN, using the ReLu activation function. Their 1D and 2D implementations achieved an accuracy of 93.5% and 94.6%, respectively, which is the highest classification accuracy identified by us. This means that both implementations, especially the 2D version, are relevant to our research study where satellite images or vegetation indices could be applied to a technique like this. More interestingly, their study found the classification accuracy for winter wheat, their only grain type, to achieve 2-3% higher accuracy than other classes. This indicates that remote sensing data of wheat grains contain phenological information relevant for distinguishing them from other crop types.

While crop classification has previously been successfully performed, there is always some uniqueness to each research effort regarding the geographical environment, target crops, and the available data. To the best of our knowledge, there has only been one previous work on crop classification in Norway, which was done by DigiFarm [22]. DigiFarm is one of Norway's leading agriculture technology companies that have developed a system for field boundaries and crop detection. An overview of their system can be seen in Figure 3.1. The figure shows their crop detection works on Sentinel images up-scaled to 1-meter resolution and the created field boundaries to classify more than six crop types. Their latest version achieves an accuracy as high as 92%. While they are using satellite images of a higher resolution than what is available in this research study, DigiFarm's systems have only been applied to specific regions in Norway. To the best of our knowledge, no research study has been performed on generalizing grain classification on all Norwegian grain producers.

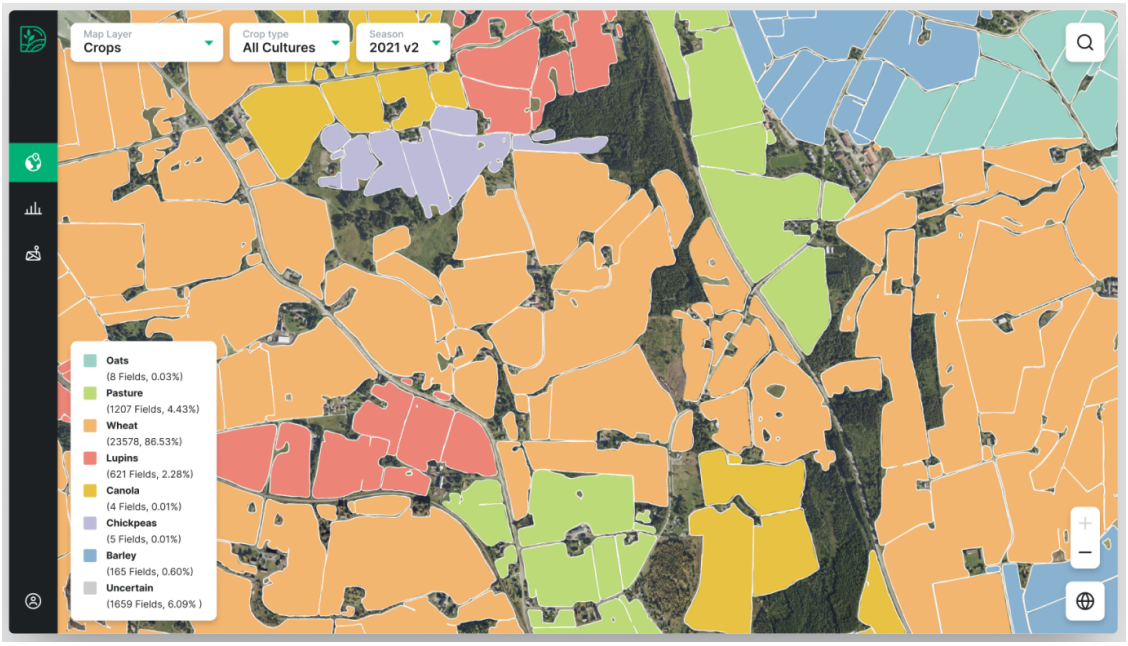

Figure 3.1: Screenshot of DigiFarm's crop detection system [22]

While the four different MIL techniques presented in Section 2.2.16 in the previous chapter improved on the state-of-the-art at their time, only the instance level classification has shown to output interpretable instance label results. However, the accuracy of this approach has later shown to be generally low [86]. This is why Ilse et al. [86] developed Attention-based Multiple Instance Learning. This technique is a single class MIL approach where a Neural Network attempts to learn the Bernoulli Distribution of the bag labels through an attention-based mechanism. This mechanism is a permutation-invariant aggregation operator and acts as an attention layer in the network. Using the weights of this layer after a prediction allows the system to output interpretable information regarding how each instance in a bag contributed to the bag's label and some indication of each instance's label. Ilse et al. [86] outperformed many previous MIL research efforts and achieved comparable results to the state-of-the-art solution. They were also able to design a more flexible MIL system that is easier to implement. Applying their system could be suitable for training a classifier with semi-labeled data for remote sensing crop classification. Ilse et al. [86] system revolve around a single target class, which means that their system would need to be extended for a multi-class environment.

3.4 The Use of Vegetation Indices for Crop Classification

Vegetation indices were shown in the previous chapter to frequently be applied to agriculture techniques and primarily crop classification and mapping. No previous research effort on crop classification limited to only grain types has been identified. This is why exploring vegetation indices and to what environment they have been applied is essential. With so many vegetation indices available, a set of indices is likely to exist that can differentiate the crop types common in Norwegian agriculture. This section will explore the state-of-the-art research on a range of vegetation indices. The research efforts included here either achieved a high classification accuracy or were able to identify some interesting characteristics of specific indices related to crop types.

Hu et al. [18] applied time series of multiple vegetation indices for a more accurate distinction of the phenological difference between corn, rice, and soybeans. The researchers used 500-m spatial resolution satellite images, which included the visual RGB bands and four different infrared bands, to create 31 sets of five vegetation indices for each image. Each set represented a day in the growing season. Based on literature and findings, the researchers chose the vegetation indices EVI, LSWI, NDSVI, Normalized Difference Tillage Index (NDTI), and Green Vegetation Index (VIgreen), and found that using multiple vegetation indices gives better results. They found the LSWI to be the most crucial feature for mapping all target classes through feature importance scores and backward feature elimination, with NDSVI being the second one. Their work shows a range of possible vegetation indices applicable for this research effort and the promising use of LSWI and NDSVI. Their work also concluded that using multiple vegetation indices over time could be extended to map other crop types in other regions, indicating that this study could be successfully applied to grain classifications.

Foerster et al. [16] used the NDVI vegetation index to classify 12 crop types. Their paper reported a highly varying accuracy between the different crop types. Winter wheat, barley, and rye obtained some of the highest accuracies, indicating that NDVI performed well in classifying grain types. The researchers also extracted how much information the NDVI index provided for each crop type throughout the growing season to evaluate which days had the most impact on mapping the crop types. Their research found that the number of images used in the growing season significantly impacted the model's ability to distinguish crops. By plotting and studying the NDVI for each crop type over a growing season, as shown in Figure 3.2, the researchers identified a trend in the continuous seasonal NDVI development. The index was identified to continuously rise during the growing season, until the crops mature, and then rapidly fall during autumn and the harvest. However, this happened at different times and values based on the crop type. Based on the figure, different periods can be identified where the index values are either similar or dissimilar between the

crop types. This type of work on evaluating the importance of each day through temporal vegetation index profiles could be used in this research effort to identify the most optimal vegetation indices or identify what days from where to include or otherwise exclude data.

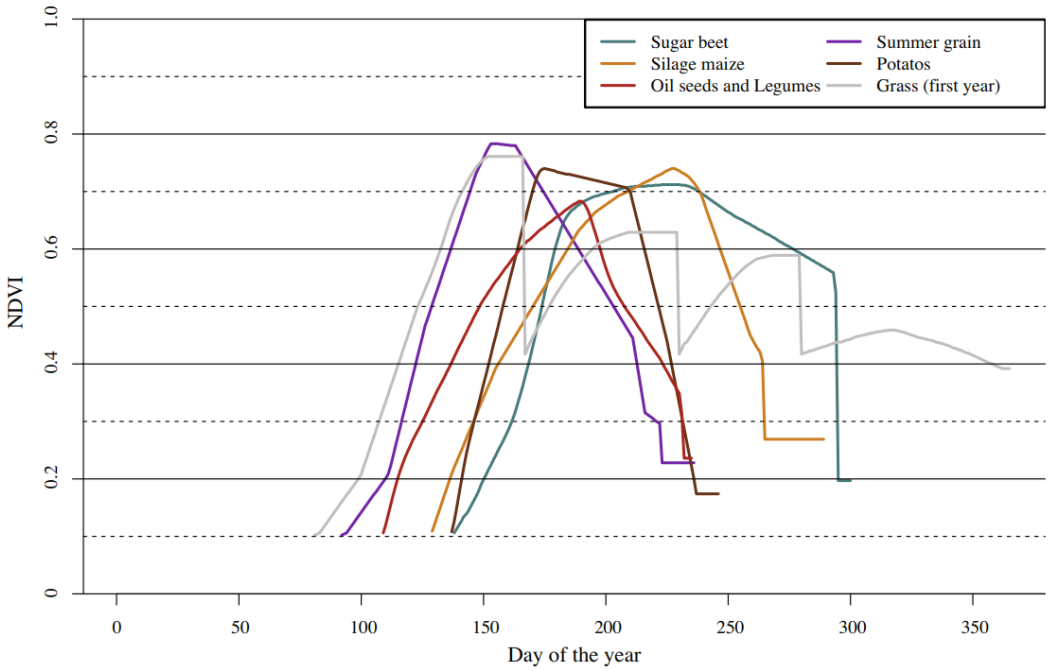

Figure 3.2: NDVI temporal profile of summer crops, by Foerster et al. [16, p. 35]

The positive effects of NDVI were also shown by Peña-Barragán et al. [82], who used several calculated vegetation indices for crop type mapping. The NDVI contributed to about 50% of the model learning in their research effort. Secondly was the SWIR-based vegetation indices, which also showed an essential contribution to the model learning. This further supports NDVI and other SWIR-based vegetation indices for crop classification.

Ma et al. [93] studied four different vegetation indices derived from multi-temporal satellite images to detect powdery mildew deceases in wheat grains. They used the Triangular Vegetation Index (TVI), Shortwave Infrared Water Stress Index (SIWSI), Disease Water Stress Index (DSWI), and Optimized Soil Adjusted Vegetation Index (OSAVI). They found the indices to be successful in predicting the wheat decease, which could further provide early information related to the health of the crops. SIWSI has also been shown to correlate with leaf and soil water content. While this indicates SIWSI to be relevant for crop health and yield prediction, it could also be suitable for crop classification through different crop types drawing different amounts of water [54].

Regarding crop leaves, Wang et al. [55] created different combinations of red, green, and blue bands and their normalized difference and researched how these indices correlated with Leaf Area Index (LAI). They found a combination of red and green, called the Green-Red Normalized Difference Vegetation Index (GRNDVI), to be one of the better when measuring LAI. GRNDVI could be a good vegetation index for differentiating crop types based on leaf sizes from the same growth stages.

Barzin et al. [56] analyzed 26 vegetation indices to predict corn yields. Their most notable vegetation indices were NDRE, GARI, SCCCI, and GNDVI, a variation of the GRNDVI presented above. These vegetation indices proved to be the most dominant variables for predicting the final yield. While their study was explicitly for corn yields, their results on the amount of helpful crop information extracted from each of the 26 vegetation indices could prove helpful for feature selection in crop classification.

Susantoro et al. [94] researched which vegetation indices to apply to map sugarcane conditions.

Their research initially started with the same thoughts as ours, where there are many vegetation indices applicable for the given problem, and a careful selection of these indices needs to happen. They analyzed 23 vegetation indices and their correlation to the plant's conditions through standard deviation and regression correlation. After some process of elimination, they concluded that NDVI, LAI, Structure Intensive Pigment Index (SIPI), Enhanced Normalized Difference Vegetation Index (ENDVI), and Generalized Difference Vegetation Index (GDVI). Most of these indices have already been presented in this chapter, or are variations of a vegetation index already presented, with SIPI being the only new one. This further strengthens the reason to apply these indices for crop classification.

Metternicht et al. [95] researched techniques within precision farming, remote sensing, and crop condition factors to develop a system for monitoring the conditions of crops and pastures. Their study revolved mainly around evaluating vegetation indices to identify field variations and condition changes. The researchers selected four indices for testing, the NDVI, Photosynthetic Vigor Ratio (PVR), Namely Plant Pigment Ratio (PPR), and the green NDVI (NDVIg). They found the PVR to be the most suitable index for distinguishing crop density variations through analysis. Their studies also found the indices to distinguish wheat crops from crops like canola and lupine. Our research effort has not previously evaluated the importance of crop density. PVR, therefore, appears as a promising index for classifying crop types, as it can apply the machine learning model with some information related to density.

3.5 Overview of State-of-the-art Research Efforts

Table 3.1 below presents a list of the relevant state-of-the-art research studies explored in our research effort, their objectives, research design, and findings. These papers will be the most significant research efforts related to our study. They will shape and inspire our research methods and be compared to our achieved results.

Authors	Objectives	Research Design	Findings
Wen et al. [91]	Minimize pre-processing and training dataset by combining cloud detection and removal	Plain and discriminative Robust Principal Component Analysis for cloud region detection and pixel restoration respectively	Developed a framework capable of detecting and removing clouds from satellite images without the need for a cloud-free image dataset
Sharma et al. [10]	Applied raw temporal satellite images for grain yield predictions	A research effort on hybrid CNN-LSTM model with multiple CNNs' to incorporate temporal remote sensing with farm information	Significantly improved the yield accuracy while also determining the significance of each growing season month and image band
Engen et al. [12, 13]	Applied satellite images, meteorological and agriculture data to neural networks to predict farm-scale yearly grain yields	An experiment and analysis on farm-based crop yield prediction in Norwegian agriculture	Achieved an MAE of 76.27kg/per decares, an improvement from their preliminary study on yield prediction without the satellite images.
Kussul et al. [14]	Applied spatial context and spectral domain information to CNNs' respectively, for crop classification	Analysis of the performance difference between pixel-based (1D) and object-based (2D) crop classification	Achieved accuracy of 93.5% and 94.6% for 1D and 2D respectively
Hu et al. [18]	Extend time-series of a single vegetation index to a time-series of multiple vegetation indices for crop classification purposes	Calculated five different vegetation indices from 7-band spatial-temporal satellite images and evaluated their performance in crop classification	Discovered LSWI and NDSVI to be the most important vegetation indices out of the five, for crop classification

Foerster et al. [16]	Crop classification and mapping using spectral-temporal and phenological features	Mapped 12 crop types from field-based spectral-temporal data and analyzed the NDVI feature throughout the season	Obtained a high accuracy for multiple crop types and saw NDVI continuously rising throughout the season with varying values for each type
Ma et al. [93]	Powdery mildew detection using multi-temporal satellite images	Derived four vegetation indices from satellite images which were applied to a k-nearest neighbors (KNN) technique for disease distribution	Found the KNN technique to provide more disease information than that from previous techniques, reaching an accuracy of 84.6%
Wang et al. [55]	Determine the best combination of red, green and blue for representing leaf area index in remote sensing	Derived all kinds of normalized difference combinations of red, green and blue band, and tested their correlation to leaf area index during different times in the growing season	Found Green NDVI, Green-Blue NDVI and Green-Red NDVI to have the highest correlation to leaf area index
Barzin et al. [56]	Develop a yield prediction model at different physiological stages	Used random forest as feature selection to determine the best vegetation indices out of 26 different, for each stage. Multiple linear regression were used for stage specific yield prediction	Determined the best vegetation indices for each stage in the growing season, and found and SCCCI to overall be the most proper one
Susantoro et al. [94]	Identified the optimal vegetation indices for classifying and mapping sugarcane conditions on fields.	Applied standard deviation analysis and regression correlation on 23 different vegetation indices, respectively. The most optimal features were then selected based on the results from both methods.	Found NDVI, ENDVI, GDVI, LAI, and SIPI to be the vegetation indices with the highest correlation to sugarcane state classification.
Metternicht et al. [95]	Researched the potential of remote sensing for monitoring the conditions of crops, pastures, and forest areas in Australia.	Examined techniques for remote sensing and precision agriculture. Identified factors for changes in crop conditions by evaluating the vegetation indices PPR, PVR, NDVI, and NDVIg.	Found the PVR to be the most optimal for crop density variation detection and the PPR to be able to detect weeds. NDVI and NDVIg were more suitable for detecting dead-standing vegetation in pastures.
Ilse et al. [86]	Solve the problem om low accuracy on instance level MIL bag classification while also having interpretable model outputs	Solved MIL problem by applying the Bernoulli distribution of the bags to a Neural Network with a permutation-invariant aggregation operator	Developed Attention-based Multiple Instance Learning, a flexible MIL solution which outperformed previous methods

Table 3.1: Overview of the state-of-the-art research efforts that related to our hypotheses and experiments

3.6 Chapter Conclusion

A wide range of research efforts that are in some form considered state-of-the-art was presented in this chapter. Initially, a method for limiting clouds in remote sensing images without using a labeled training set was introduced. Crop yield prediction saw two studies highly relatable to our research effort. Sharma et al. [10] presented promising techniques for predicting yield using high spectral-temporal data. Engen et al. [12, 13] were presented as the yield prediction research effort most similar to ours, while Digifarm [22] was presented as the most similar classification study. Besides this, Kussul et al. [14] were presented as the ones who achieved the highest crop classification accuracy (94.6%). Lastly, a wide range of vegetation indices was presented through successful research efforts. The most noticeable indices applicable to crop classification were NDVI, NDSVI, EVI, LSWI, SWIR, and SIPI. Most fascinating was the exploration of the temporal development of NDVI, which will be discussed in forthcoming chapters.

Chapter 4

Data Acquisition and Pre-Processing

A wide range of different spatial, spectral, temporal, and textual features are applied throughout this research effort to different AI models, including remote sensing, geographical, meteorological, and grain production data. Therefore, it is essential to present all of our features, what information they contain, and their sources. A substantial part of the methods in this research effort is related to data pre-processing. This chapter describes how the data mentioned above was collected and pre-processed before being used for machine learning. The chapter accounts for data pre-processing required for each of the experiments that will be performed in Chapter 5. Firstly, we describe the remotely sensed data and how masks were applied. Further, the geographical data will be presented, followed by all temporal meteorological data. Lastly, we explain how grain production and farmer grants were collected and pre-processed.

4.1 Spectral-Temporal Remote Sensing Images

This section will cover the remote sensing data applied in this research effort, how the data was acquired, the structure, and the information. This will consist of satellite images and vegetation indices. Further, we will describe how images were cropped and why, while visualizing examples. For visualization purposes, we use the same farms and fields for all our examples to better understand the resulting changes of our techniques

4.1.1 Sentinel-2A Satellite Images

Satellite images are remotely sensed data collected by different satellites orbiting the earth. These images are openly available globally and include detailed high-resolution observations of the planet. An extensive set of Sentinel-2A satellite images of Norwegian farms were delivered by KORNMO. These images were centered on a farm and reflects the surrounding area of that farm. These raw satellite images are therefore also referred to as farm-based satellite images. The acquired images contain 12 different spectral bands with different meanings, wavelengths, bandwidths, and resolutions, as shown in Table 4.1. The images were combined and sorted in a file type called H5. An H5 file is a data file saved in a hierarchical format that can store multidimensional arrays. This file type allowed us to look up the photos for each farm for each year, where every farm has 30 images. The satellite images consist of data from 2017 to 2019. The API codebase used by Engen et al. [12, 13] to extract satellite images from Sentinel Hub[1] were re-implemented. Satellite images for 2020 were attempted to be downloaded but were incomplete because of financial limitations. This will be addressed more in the discussion i.e., in Chapter 7. The acquired images were taken

[1] https://www.sentinel-hub.com/

from the 1st of March to the 1st of October to match Norway's growing season. The previous work done by Engen et al. [12, 13] states that this period yields the best results for yield prediction. Figure 4.1 shows an example of all 30 raw satellite images of a farmer in a given year. Only the Red, Green, and Blue bands are depicted for visualization purposes, as the other nine bands are not that suitable.

Band	Name	Wavelength	Bandwidth	Resolution
1	Ultra Blue Coastal Aerosol	442.7	21	60
2	Blue	492.4	66	10
3	Green	559.8	36	10
4	Red	664.6	31	10
5	Vegetation Red Edge	704.1	15	20
6	Visible and Near Infra-red	740.5	15	20
7	Visible and Near Infra-red	782.8	20	20
8	Visible and Near Infra-red	832.8	106	10
8a	Visible and Near Infra-red	864.7	21	20
9	Water Vapor	945.1	20	60
11	Short Wave Infra-red	1613.7	91	20
12	Short Wave Infra-red	2202.4	175	20

Table 4.1: Overview of Sentinel-2A bands, their specifications, and the information they reflect [96].

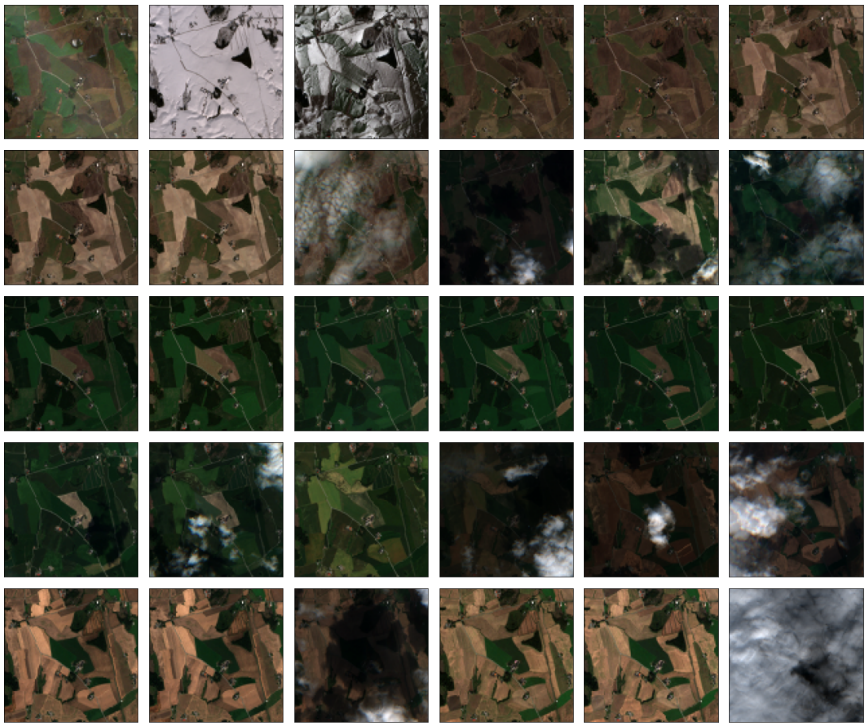

Figure 4.1: An example of the 30 raw satellite images throughout a growing season of a specific farmer, visualized in RGB

4.1.2 Masking Remote Sensing Images

In order to filter out pixels and information from remote sensing images, masks were applied. Masks for the original farm-based images were acquired through KORNMO and hold the same size as the images but with one channel ($100\ x\ 100$). Additional masks were created throughout this research effort for different types of remote sensing images. A generalized algorithm of this process is given in Algorithm 1. Masks were created based on a farm's geometry in *disposed properties* (Explained

later in this chapter), reflecting the farm's arable land area. The farmer's coordinates are used as the center of the mask to exactly match the satellite images. Each pixel in the masks is either *True* or *False*, where a pixel value of *True* means that the pixel is within a property boundary belonging to the farm. Before remote sensing images were used, each channel was multiplied by the mask to convert non-property pixels to zero, removing unnecessary information for the models. This process is visualized in Figure 4.2, where the left side shows four satellite images of the same farm, and the right side shows the same images with masks applied.

Data: target dimension (NxN) and property shape
Result: Masks of dimension NxN, centered on a property shape
initialization;
for *farm in all_farms* **do**
  for *year in farm_years* **do**
    get property shape;
    get property center coordinates;
    created NxN dimension bounding box;
    convert property shape to polygon;
    for *point in polygon_coordinates* **do**
      translate coordinates to xy values;
    end
    draw image using xy values;
    fill empty values in image;
    store mask in h5-file;
  end
end

Algorithm 1: Algorithm for creating NxN sized masks for filtering out non-property data in images.

Multiple types of remote sensing data were applied throughout this research effort, and masks were created for each of these unique types. These will be explained and visualized throughout this section when their respective remote sensing data are presented.

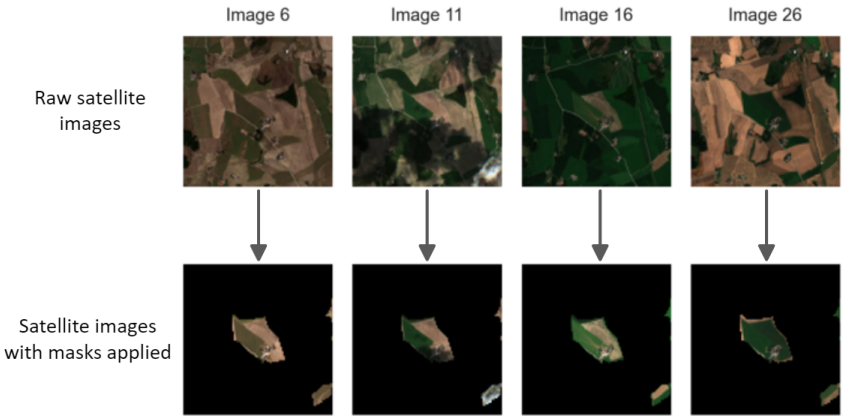

Figure 4.2: An example of property masking applied to four 100 x 100 satellite image of a farm, spread throughout the growing season

4.1.3 Field-based Remote Sensing Images

While the original (100 x 100) satellite images were applied to classification models for crop identification as an initial experiment, which will be explained in Chapter 5, they were not practical to use. The original satellite images contain multiple fields, often having different crop types planted

between them. A single label can therefore not be applied to every image. In addition, we do not want the classification model to consider or learn the location of a field concerning its farm or center (coordinates), as this is unrelated to the planted crop type. A crop classification model should be fed information about a field or an area containing only one crop type, so the crop type of future fields can be predicted without access to the labels. While object identification and classification might be suitable for satellite images of a complete farm to map multiple crop types in one image, it is likely to be difficult without accurate boundaries between each field. Field-based images were therefore created for every field of each farmer.

Fields are significantly smaller than a farmer's property, and a field-based dataset is likely to contain more than five times as many images as a farm-based dataset. Therefore, the new field-based images were created with a size of (16 x 16), which covered only one field each, centered on the field center. This exact resolution was selected after some visualization and testing, which showed the resolution to cover most fields while still being significantly smaller than the larger images. To create these images, we used the original satellite images, positional data of the farmers, and the field-based classification dataset (Section 4.2.4). The algorithm's pseudo-code for creating field-based images is shown below in Algorithm 2. For each farmer and field, we translated the field's latitude and longitude to a value between 0 and 100, corresponding to its center position on the original satellite image. If the edges of the field's position were outside the satellite image, it was shifted by an offset. When the field is placed correctly, we cropped the satellite image using a size of (16 x 16). An example of these field-based images is shown on the left side of Figure 4.3, showing four fields throughout the growing season.

A field mask was created using the same shape for each field image to filter out non-field information. This masking is shown on the right side of Figure 4.3. The new field-based images and their respective masks were stored in different h5-file using the same structure as the initial h5-file.

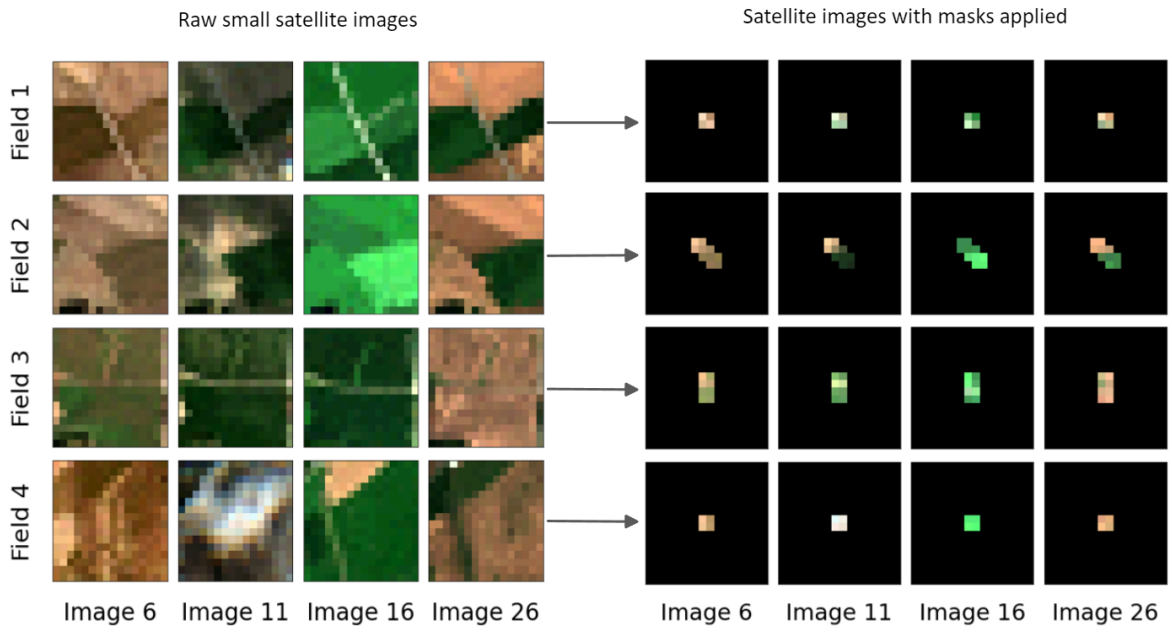

Figure 4.3: **Left side:** Four examples of four 16 x 16 fields throughout the growing season, created by the original satellite images in Figure 4.1. **Right side:** The field-based images from the left figure with field masks applied.

```
Data: 100x100 satellite data, positional data, classification data
Result: 16x16x12 satellite images
initialization;
for farm_images in all_images do
    polygons ←polygons for current farm;
    for polygon in polygons do
        translate positional data;
        for image in farm_images do
            crop image;
            store cropped image in h5-file;
        end
        create mask from polygon;
        crop mask;
        store cropped mask in h5-file;
    end
end
```

Algorithm 2: Algorithm for creating 16 x 16 images

4.1.4 Vegetation Indices Derived from Satellite Images

Throughout the previous research efforts explained in Chapter 3, the promising use of vegetation indices for different purposes was shown. In particular, a range of research efforts presented by Hu et al. [18], Ma et al. [93], Wang et al. [55], Barzin et al. [56], showed how well different vegetation indices were in classifying and mapping crop types. A set of vegetation indices were therefore derived from the raw sentinel images. Based on the findings of each of the research efforts, a set of promising indices that appeared suitable for Norwegian agriculture were selected. The correct Sentinel-2A bands for each vegetation indices were identified from Table 4.1, based on the wavelength information from the theory in Chapter 2, Kobayashi et al. [50] and Henrich et al. [37]. Based on the previously presented theory, the following formulas were used to create vegetation indices from every available raw satellite image. The band numbers in each formula are the same as the Sentinel-2A bands in Table 4.1.

EVI

EVI were calculated based on Equation 2.1 which uses Blue, Red and Visible and Near Infra-red (VNIR) bands. The tunable values have been selected for Sentinel-2A specific environment and settings:

$$EVI = G * \frac{band8 - band4}{(band8 + C_1 * band4 - C_2 * band2) + L)} \tag{4.1}$$

$$\text{G} = 2.5 \quad C_1 = 6 \quad C_2 = 7.5 \quad \text{L} = 1$$

LSWI

LSWI were calculated based on Equation 2.2 which uses VNIR and Short Wave Infra-red (SWIR) bands:

$$LSWI = \frac{band8 - band11}{band8 + band11} \tag{4.2}$$

NDSVI

NDSVI were calculated based on Equation 2.3 which uses Red and SWIR bands:

$$NDSVI = \frac{band11 - band4}{band11 + band4} \tag{4.3}$$

NDVI

NDVI were calculated based on Equation 2.4 which uses Red and VNIR bands:

$$NDVI = \frac{band8 - band4}{band8 + band4} \tag{4.4}$$

SIWSI

SIWSI were calculated based on Equation 2.5 which uses Red and VNIR bands:

$$SIWSI = \frac{band8a - band11}{band8a + band11} \tag{4.5}$$

GRNDVI

GRNDVI were calculated based on Equation 2.6 which uses Green, Vegetation Red Edge and VNIR bands:

$$GRNDVI = \frac{band8 - (band3 + band5)}{band8 + band3 + band5} \tag{4.6}$$

NDRE

NDRE were calculated based on Equation 2.7 which uses Vegetation Red Edge and VNIR bands:

$$NDRE = \frac{band7 - band5}{band7 + band5} \tag{4.7}$$

SIPI3

SIPI3 were calculated based on Equation 2.8 which uses Blue and VNIR bands:

$$SIPI3 = \frac{band8 - band2}{band8 - band4} \tag{4.8}$$

PVR

PVR were calculated based on Equation 2.9 which uses Green and Red bands:

$$PVR = \frac{band3 - band4}{band3 + band4} \tag{4.9}$$

GARI

GARI were calculated based on Equation 2.10 which uses Blue, Green, Red and VNIR bands:

$$GARI = \frac{band8 - (band3 - (band2 - band4))}{band8 - (band3 + (band2 - band4))} \tag{4.10}$$

These vegetation indices were calculated by extracting the relevant bands from the already acquired Sentinel-2A satellite images and combining the bands based on their formula. These ten vegetation indices were calculated for each of the 30 satellite images in the growing season for each farm and stacked together into a [30 x 100 x 100 x 10] image. Four examples of each vegetation indices are shown on the left side of Figure 4.4, with their property mask applied on the right side. The images were normalized between 0 and 1 to visualize the differences within the image. The images are visualized in grayscale, as they only have one channel. The reason why some of the images are blurred or close to blank can be several reasons, such as clouds or other noise, limitation to the vegetation index, or limitations in the visualization.

In order to apply the vegetation indices for classification, a field-based dataset was created from the vegetation indices images, just as it was done for the raw satellite images in Section 4.1.3. Four examples of two fields with the vegetation indices NDVI, LSWI, and NDSVI are shown on the left side of Figure 4.5, with the same field-based images with field masks applied on the right side.

4.2 Geographical Data

Geographical data is essential in this project for extracting and creating accurate features. Most importantly, this data was used to create the shape and boundaries of fields. This helped the farmers to remove irrelevant data and map the soil quality. The geographical data was also used to create the smaller, field-based images, features for the yield prediction model, and data exploration when visualizing fields and farms on a map. This section will describe all of the geographical data used in our research efforts, such as properties, coordinates, and soil quality. It also describes how some of it was created and applied in our research effort.

4.2.1 Disposed Properties

The disposed properties dataset was acquired from the previous work done by Engen et al. [12, 13]. It was created by an intersection between the Norwegian cadastral data filtered by grain farmers and the NIBIO soil quality dataset. This intersection results in boundary property data only including arable farmland excluding trees, rivers, and rocks. This dataset now contains geometry polygons of field boundaries, grouped into a multi-polygon, for each farmer, from 2017 to 2019. This dataset creation is further explained in work done by Engen et al. [12, 13], but an example is given in Figure 4.6. The left image shows the geometry polygons in disposed properties plotted on a spatial view of the earth. The right image is of the same area but zoomed in on the farmers whose remote sensing data were presented in the previous section. The boundaries of the arable farmland area are clearly shown for this farmer. It is worth noticing that a geometrical shape in disposed properties might consist of multiple actual fields. The disposed properties dataset is used to connect the soil quality dataset to each farmer, which allows the usage of soil quality in yield prediction, and to create a field-based dataset with their planted crop types for crop type classification.

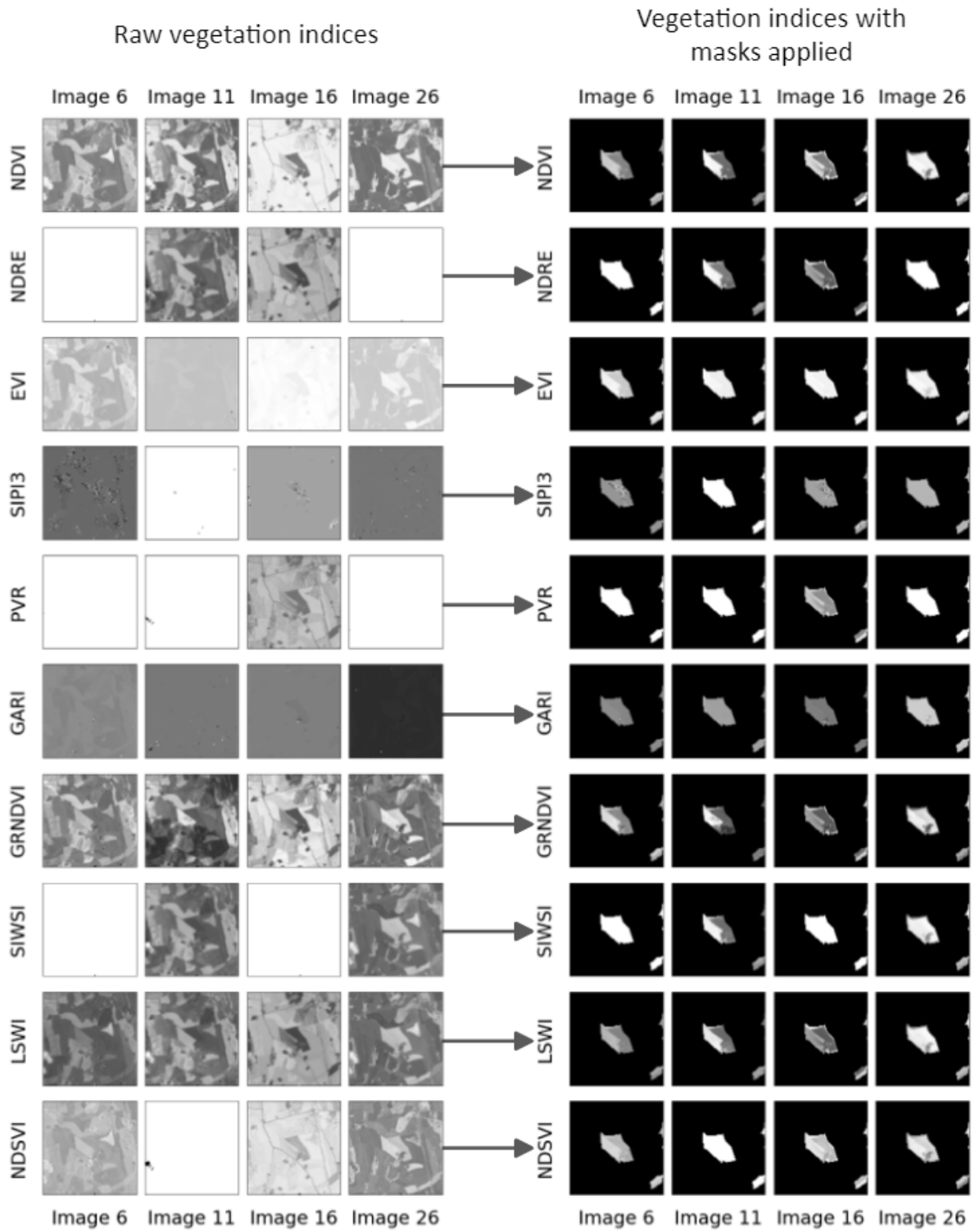

Figure 4.4: **Left image:** Ten 100 x 100 vegetation indices derived from the satellite images in Figure 4.1, spread throughout the growing season. **Right image:** The vegetation indices from the left figure with property masks applied.

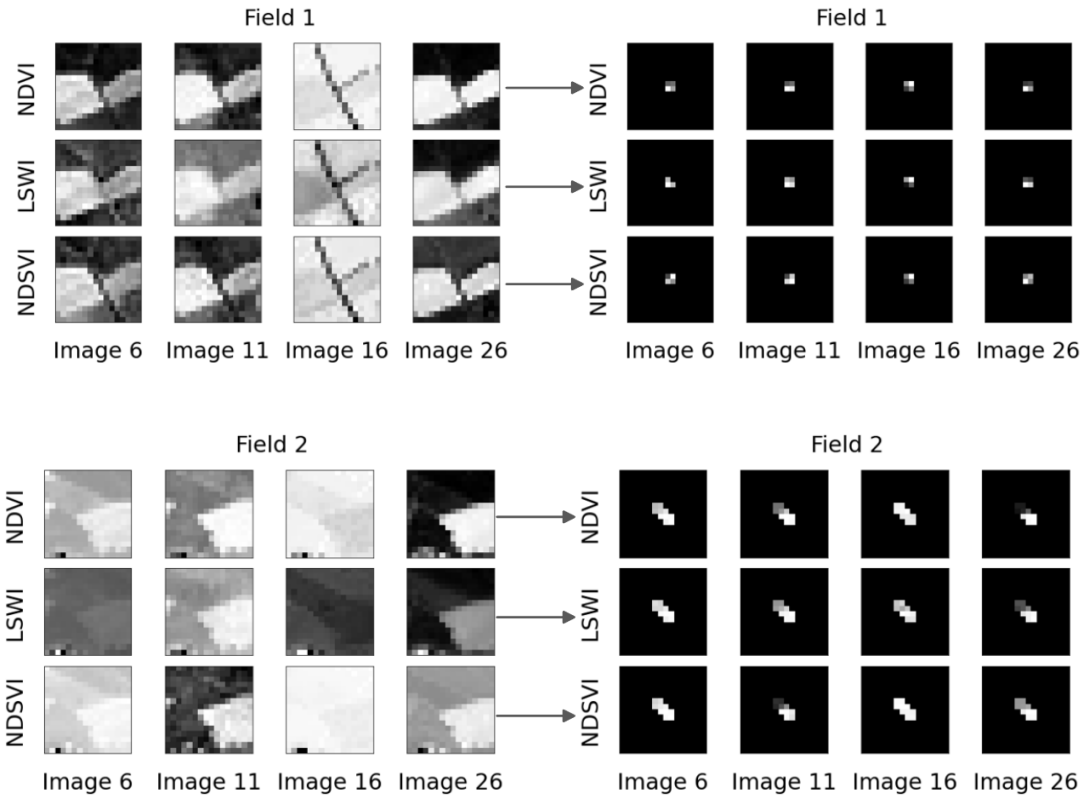

Figure 4.5: **Left side:** Two sets of 16 x 16 field images throughout the growing season, for the vegetation indices NDVI, LSWI and NDSVI, created by the vegetation images in Figure 4.4. **Right side:** The field-based vegetation indices from the left figure with field masks applied.

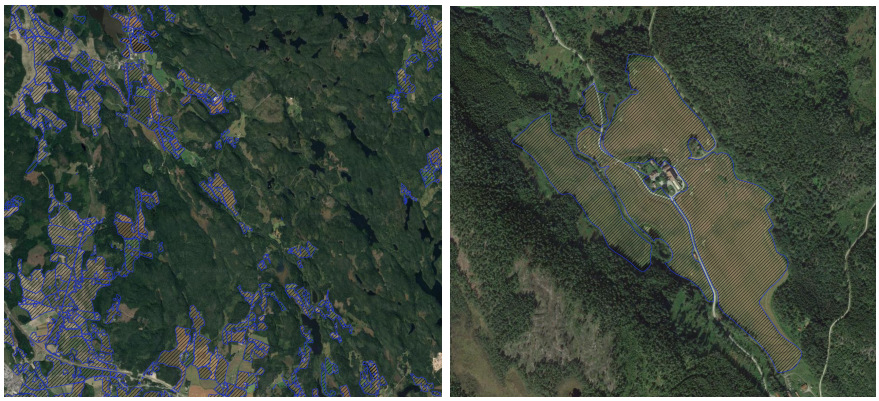

Figure 4.6: Disposed properties visualized for a large area (left) and for a specific farmer (right)

4.2.2 Coordinates

In order to download new satellite images, the location of each farmer had to be acquired. This is easiest handled in the Sentinel system by using coordinates. The geographical location is also used as features for yield prediction, as the location where a farmer resides in Norway can impact grain production. Factors such as altitude, coastline, and the general geography of the surrounding area can impact the production both ways.

Address and location are publicly managed by The Brønnøysund Register Center (Brønnøysundregistrene), along with a wide range of other information related to enterprises. This also includes the address for farms, as farms in Norway are registered as a type of business enterprise. The Brønnøysund Register Center develops and operates 18 different digital register services, making public data accessible for individuals and companies. Their most important registers are related to enterprise and business information, bankruptcy registers, and marriage agreements. Their goal is to handle public data sources, providing data with safety and order while contributing to high-value creation for actors [97].

Using the organization number of farmers, which will be explained in Section 4.4 later in this chapter, each farmer's municipal, postal code, and address were acquired. This was done by querying the register center's Central Coordinating Register for Legal Entities API[2], the register related to business enterprises.

Mapbox designs precise location data and powerful tools for developers, focusing on high-performance, distributed services accessible on low-powered devices with low bandwidth worldwide [98]. While latitude, longitude, and elevation are available through various services and agencies, Mapbox provides some of the simplest ways of acquiring them. They manage a Geocoding API[3] which performs forward and reverse geocoding. The forward geocoding converts an address into latitude and longitude coordinates, while the reverse geocoding does the opposite. The farmer's address information acquired from the Central Coordinating Register for Legal Entities API above was used to query Mapbox's Geocoding API to acquire the latitude and longitude of each farm. In addition, Mapbox manages a Tilequery API[4], which handles vector tilesets based on a given latitude and longitude. The latitude and longitude for each farmer acquired from the previous Mapbox API were used to query the Tilequery API to acquire the elevations of each farmer.

After some evaluation of the coordinates, we identified a significant number of farms whose organization number were registered to a different geographical location compared to the actual location of the farm. This meant that the registered address belonged to a different property, sometimes far away from the farm, within a city. Extracting satellite images based on these coordinates would result in miss leading information. Therefore, new latitude and longitude values were calculated from the geometry polygons of a farm property in *disposed properties*, from Section 4.2.1. Mapbox's Tilequery API was then queried again for more accurate elevation values using the new latitude and longitude values. These new coordinates were further identified to be significantly more accurate than the previous values, having close to double the number of decimals, narrowing the geographical location.

4.2.3 Soil Quality

The Norwegian Institute of Bioeconomy Research (NIBIO) holds and maintains the field geographical locations. In addition, they organize soil quality dataset with the purpose of documenting Norwegian food soils' properties as a resource. The dataset contains 5300 square kilometers of

[2] https://data.brreg.no/fullmakt/docs/index.html
[3] https://docs.mapbox.com/api/search/geocoding/
[4] https://docs.mapbox.com/api/maps/tilequery/

Norwegian agricultural fields, collected over 40 years. Each agricultural field in the dataset contains information about the quality of the soil and landscape that can provide information about the field level, what the soil contains, the amount of rocks, how water runs through the soil, what is suitable for planting on the specific field, and much additional information. The dataset can be accessed on NIBIO's site [99] along with a more detailed description of the information.

The raw dataset contains 78 features, and many of these are either duplicates, only have one possible value, or are information irrelevant to grain production. One of the features is a geometry feature containing the geometry polygon of the relevant area. An example of this is shown in Figure 4.7. The left image presents different levels of soil quality and their geometry in a large area, and the right image presents the soil quality of different areas of a specific farmer. Overall, the soil quality dataset features contribute to the complex calculation of a single main feature, called soil quality. Soil quality represents the overall quality of the soil as one out of three values:

- 1 = Very good quality, shown as blue in the figure
- 2 = Good quality, shown as green in the figure
- 3 = Less good quality, shown as red in the figure

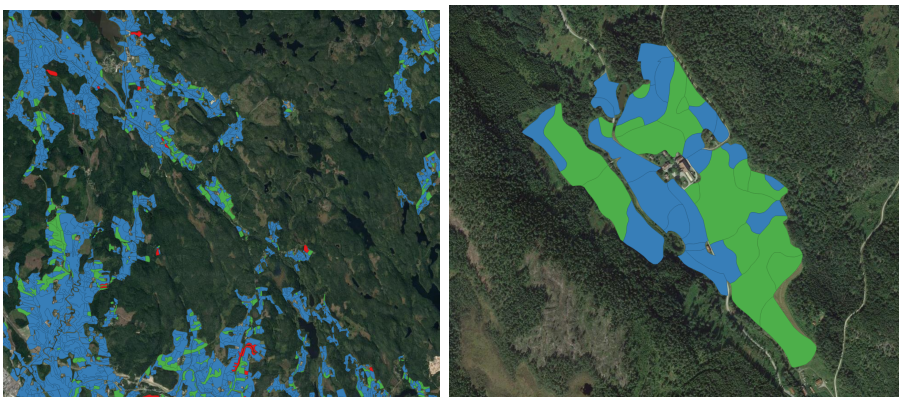

Figure 4.7: Soil quality visualized for a large area (left) and for a specific farmer (right). The blue, green and red colors reflects soil quality 1, 2 and 3 respectively.

It is worth noticing that an actual field and a geometry in *disposed properties* might contain multiple soil quality entries, with different values explaining the field. NIBIO creates the soil quality feature to represent how the fields are to work with and maintain and how strong a yield they can produce. It also represents information such as the field's incline, the amount of rocks, and the distribution of element particles. This feature can indicate how grain crops will grow and mature in the given soil and are therefore used as an average quality for the farmer during the yield prediction.

In order to be able to map the soil quality data to each farmer, every shape in the soil quality dataset needed to be identified and mapped to a specific farmer. A python script that used the farm boundary dataset from Section 4.2.1 and the soil quality dataset was developed to accomplish this. The python code checked what field in the soil quality dataset intersected which farm boundary. Then for each intersection, the field id, the municipal and organization number of the farm were appended to the new dataset. This new dataset served as a lookup table between each field and their belonging farm and is how soil quality was connected to each farm. The lookup table contained a high amount of fields for each farm. In order to enable to use of soil quality for yield prediction, the lookup table was used to create an average soil quality dataset, limiting the number of features for each farmer. Each average soil quality dataset entry contained the mean soil quality and the number of fields with a farmer's soil quality value of 1, 2, and 3.

4.2.4 Field-based Classification Dataset

A new dataset containing what each farmer planted each year was made to match fields with the satellite images. This dataset would filter what images the machine learning model should use based on organization number and year. This dataset was created based on one assumption: if each farmer only delivered one type of grain the entire year, it was assumed that all of the farmer's fields only contained that grain type. The types of crops delivered in grain deliveries and properties geometry from *disposed properties* were used to create this dataset. Grain deliveries were used to find which farmers delivered one kind of grain each year. Figure 4.8 shows some example rows in our delivery data, where the first row would be included in the dataset with a label of *bygg* combined with organization number and year. The second row would not be part of the dataset and would instead be added to the unlabeled data used in our Multiple Instance Learning model.

year	orgnr	komnr	bygg_sum	erter_sum	havre_sum	hvete_sum	rug_sum	rughvete_sum	oljefro_sum
2019	811935662	3014	45513	0	0	0	0	0	0
2019	813002302	704	0	0	0	202481	68115	0	0
2019	813865122	3419	0	0	78832	0	0	0	0
2019	814893332	621	0	0	33547	0	0	0	0

Figure 4.8: An example of the deliveries dataset

In addition, the field boundaries were added to the dataset in order to perform field-based classification using the field-based images. The field boundary data is included in the *disposed properties* dataset. Each farm for each year has a multi-polygon data structure that consists of all boundary data for that farm. The multi-polygon was split up and matched with what was planted so that each row in the new dataset only had one polygon (one field) combined with the organization number, year, area, and what was planted. Figure 4.9 shows the some example entries and how the dataset was structured.

year	orgnr	planted	area	geometry
2019	811935662	bygg	94540.480524	POLYGON ((301788.79090 6608202.23000, 3...
2019	811935662	bygg	22778.560014	POLYGON ((301588.81297 6607862.03284, 3...
2019	811935662	bygg	5072.924216	POLYGON ((301544.54510 6608141.21000, 3...
2019	811935662	bygg	11611.972746	POLYGON ((301285.90120 6608685.32060, 3...
2019	812530542	havre	10203.282317	POLYGON ((281369.42428 6687420.41295, 2...

Figure 4.9: An example of the field-based classification dataset structure

4.3 Temporal Meteorological Data

Meteorological data are first and foremost important for the farmers. Agriculture and plant growth requires the right conditions at the correct times for crops to grow successfully. These factors are essential for achieving satisfactory yield at the end of the growing season.

The old and the new meteorological features are extracted through the Frost API, as done by Engen et al. [12, 13]. The Frost API[5] is created by The Norwegian Meteorological Institute, which collects meteorological data from more than 1319 stations across the whole country. As of this writing, the API manages 658 different features covering everything from temperatures, precipitation, solar factors, wind, pressure, and more. Using knowledge attained through the preliminary project by Engen et al. [12, 13], the theoretical background in Chapter 2, and previous research efforts in Chapter 3, all 658 features were evaluated to identify new features suitable for crop yield prediction.

[5] https://frost.met.no/index.html

While the Frost API contained multiple features relevant to crop growth, many of them, such as solar radiation and other solar factors, were only recorded by a few measurement stations. The features that were available countrywide, which were also identified to reflect some information related to crop growth, were the following:

[Growth Degree, Sunlight Duration and State of the Ground]

4.3.1 Meteorological Features

Temperature and Precipitation

Temperature and precipitation have already been shown to be relevant for plant growth in Norwegian agriculture through the study by Engen et al. [12, 13]. In addition, VanDerZanden [26] and Johnson et al. [27] showed that temperature and precipitation are some of the most critical factors for crop growth, where a correct combination of the two is needed for plants to thrive. The precipitation is extracted from the Frost API as one value for each day, for each station, representing the total accumulated precipitation for that day, measured in millimeters. The temperature is extracted as three values for each day. The three values represent the lowest, highest, and daily average air temperature, listed in degrees Celsius and were measured two meters above the ground.

Growth Degree

Growth degree in the Frost API, also known as the sum of the heat during a whole day, measures how good the conditions are for plant growth. This heat sum is defined as the number of degrees the daily average temperature is above a particularly given value, which here is 5 °C. Every value equal to 5 °C or less is set as zero. While growth degree is a form of average temperature, it presents information more related to the conditions for plants to grow. Treating average temperatures less than six the same might enable a machine learning model to learn more from the values above five, which have more impact on the plant's growth. Growth degree can be managed through different formulas, such as the average daily temperature multiplied by the number of days in the growing season [25]. Nevertheless, generally, growth degree gives information about how much heat the crop has been exposed to for growth above a threshold. These are the reasons for applying growth degrees to our experiments [25]. The growth degree is extracted from the Frost API as one value for each day, for each station, listed in degrees Celsius. A set of growth degrees' were also given to us through actors in KORNMO, which further strengthened the reason for applying it, as it was already available in some size.

Sunlight Duration

Sunlight has also been shown to be essential for crops in general and specifically crucial for grains. Fields and areas with the most sunlight are often used to grow grains and can tackle lower temperatures [25]. A certain amount and type of light are also needed to transition grain seeds into seedlings, all the way to mature plants. Sunlight duration is extracted from the Frost API as one value for each day, for each station, which contains how many hours of sunlight the sensor has experienced during the last 24 hours, listed as hours.

State of the Ground

Temperature, precipitation, growth degree, and sunlight duration have to this point, all been shown to relate to crop growth in some way. The state of the ground has the potential to represent these features in a temporal context. A cracked, dry ground is likely an indication of a high temperature and minimal rainfall in the last period. A thick layer of snow indicates that the area has experienced rainfall while temperatures were below zero in the last period. This is why the state of the ground is interesting for crop yield, as the state of the field can say a lot about the climate over time. VanDerZanden [26] also showed how the state of the fields is essential for the farmer, requiring dry soil for their equipment.

The state of the ground is a feature containing information about the ground and the state of the soil. Each station records the ground state four times each day, with a six-hour interval (00:00, 06:00, 12:00, 18:00) where the state is logged as a number, representing a specific class. This recorded measurement mainly yields information regarding the current state of the soil and its surface, how moist it is, and how much water in some form is covering the ground. The selected regions in Norway contained measurements covering the 17 unique classes shown in Table 4.2.

Class Value	State Description
-1	No snow or ice cover
0	Dry surface
1	Moist ground
2	Wet ground
3	Flooded ground without snow
4	Frozen ground with measurable ice
5	Glaze on ground
10	Ground covered by ice
11	Compact of wet snow covering less than half the ground
12	Compact or wet snow covering more than half the ground
13	Even layer of compact or wet snow covering ground completely
14	Uneven layer of compact or wet snow covering ground completely
15	Loose dry snow covering less than half the ground
16	Loose dry snow covering more than half the ground
17	Even layer of loose dry snow covering ground completely
18	Uneven layer of loose dry snow covering ground completely
19	Ground Completely Covered by Snow

Table 4.2: Descriptions of the original states of the ground and their class number [100]

The meteorological features explained in this section are stored and handled by the Frost API with the following IDs:

- **Temperature** - air_temperature
- **Precipitation** - sum(precipitation_amount P1D)
- **Growth Degree** - integral_of_excess(mean(air_temperature P1D) P1D 5.0)
- **Sunlight Duration** - sum(duration_of_sunshine P1D)
- **State of the Ground** - state_of_ground

4.3.2 Feature Pre-Processing

The API scripts created in the recent works by Engen et al. [12, 13] were used to download and process data from the 2020 growing season. The scripts were then customized to handle the three new features; *Sunlight duration*, *State of the Ground*, and *Growth degree*. The scrips were also modified to be more user-friendly and adaptable for new features. The researchers explain the codebase and processing of the features in [12, 13], but a summarization is given below.

1. Downloads all available Frost measurement stations with geographic information.
2. Downloads raw measurement readings for each day from each sensor above, using the measurement station ID and feature ID as payload.
3. Extracts the measurement value from the raw readings and processes the measurements as time series with one value for each day in the growing season for each station.
4. Validates each feature dataset, a new step created in this research effort. Every sensor which has less than 30% (65 days) of the measurements for a season is dropped. Any missing values are then replaced using linear regression based on the distribution of all values for that sensor. This process ensures that there are zero Nan values in the meteorological datasets while also providing as much information as possible.
5. Further, one of the following is applied:
    - For each feature, a neural network is trained to interpolate all measurement values by learning the relation between distance and changes in measurement values. This enables the model to predict a local measurement for a specific farmer based on its three closest measurement stations. During training, the neural network attempts to predict the measurement value of a station based on the distance and value to its three closest stations. The model is able to learn through this method, as the actual measurement value of the target station is already known. After successful training, the model attempts to predict the measurement value of a farmer's location based on its three closest stations.
    - *Ground State* was unavailable for measurement interpolation, as the trained Neural Network achieved an unpractical high loss and could not learn the relation between measurement value and distance. In situations like this, a more straightforward proximity-based filler is used. This method fills all measurement values of a farmer with the measurement of its closest station.

As mentioned above, the scripts were customized to handle new features. The *Sunlight duration* and *Growth degree* feature were of an almost identical structure to that of the already implemented *Precipitation*, having a single value for each day. Therefore, these two new features were implemented in the scripts based on the existing codebase.

The state of the ground was trickier to implement as each day contained four measurements spread out over the day. Adding these features (860) would probably be too much for our data distribution and AI models. Further, each measurement represented a state of the ground, which meant that averaging the values into a daily value would not be correct. To solve this problem, the scripts regarding the third bullet point in the list above were customized to perform specific actions for the *State of the Ground* feature, to end up with one daily value eventually. The different state classes were narrowed from 17 to 11 by merging similar states and converting the class value. The merging of similar states, shown in Figure 4.3, made it possible to find many days with the same states for each of the four measurements.

The following algorithm ensured that each sensor's measurements collected over a day were merged into one daily value. Each day might not always have four measurements, as the sensors cannot

New Class Value	State Description	Original Class Values
0	Dry surface	0
1	Moist ground	1
2	Wet ground	2
3	Flooded ground without snow	3
4	Frozen ground with measurable ice	4
5	Glaze on ground	5
6	Ground covered by ice	10
7	Snow covering less than half the ground	11, 15
8	Snow covering more than half the ground	12, 16
9	Snow covering ground completely	13, 14, 17, 18, 19
10	No snow or ice cover	-1

Table 4.3: Overview of the new ground state classes and which of the original classes they cover.

always record a state. The n, therefore, represents the number of raw measurements for that day. This algorithm enables each station to receive a daily value which is the most accurate at representing the four measurements throughout a day.

1. If one class value is represented at least *n-1* times, this is used as the daily value.

2. Else, if it is 50/50 between two classes, the following conditions apply:

   - If the two classes are *Moist* and *Wet ground* $->$ *Wet ground*.
   - If the two classes are a mix of any amount of snow coverage $->$ *Snow covering more than half the ground*.
   - If the two classes are *Glaze* and *Ground covered by ice* $->$ G*round covered by ice*.

3. If one class has at least two occurrences, then use this.

4. If a daily value is still not found, then print this value and manually solve it.

4.4 Norwegian Grain Production Data

The Norwegian Agriculture Agency[6] currently collects and maintains the most detailed and important records related to agriculture and farmers' activities. Two noticeable data sources are the deliveries to grain buyers or seed shops in agriculture, hereby named *grain deliveries*, and the production and replacement subsidies for agricultural enterprises, hereby named *production subsidies*. These two data sources are publicly available from The National Data Catalog at https://data.norge.no/ and are published as datasets for each growing season. The records in these data sources contain vast information about each farmer in a given growing season related to fieldwork, production, and husbandry, among others. Each farmer in both data sources is listed with an organization number unique to that farmer. This organization number connects farmers' records with records from other information providers such as Kartverket[7] and Brønnøysundregistrene[8]. The organization number is frequently used throughout this research effort and may also be referred to as *farmer ID*.

[6] https://www.landbruksdirektoratet.no/nb
[7] https://www.kartverket.no/en
[8] https://www.brreg.no/

4.4.1 Grain Deliveries

The grain deliveries dataset contains the amount of grains and seeds in Kg each farmer delivered to grain buyers or seed shops in agriculture. The dataset is an important but simple dataset covering deliveries from July in the given growing season until June the following year. It is usually published in October of that year. The dataset contains 18 features that separate grains between different qualities and purposes for the following grain types:

[Barley, Wheat, Oat, Rye, Rye Wheat, Oil-seeds, Pea]

Grains are delivered as either animal feed, sowing seeds, or products for human consumption, which is often determined by the quality of the grains, whereas grains for human consumption have the highest quality. The previous work by Engen et al. [12, 13] chose to exclude oil-seeds and pea as they are not grain crops. These will, therefore, also not be applied for the grain yield predictions in this research effort.

Grain deliveries were extracted from The National Data Catalog[9], first manually for exploration purposes and then automatically in our yield prediction system through their API. The information was extracted for every growing season from 2012 until 2021, where the earliest data were applied as supplementary information representing historical data. As grain deliveries contain the amount of grains each farmer delivers in a growing season, it can also be considered the total actual yield for a farmer. While the dataset might not cover 100% of what every grain farmer in Norway produces, as it does not record information regarding rented ground grain or rented cleaned seeds, it still serves as the most accurate information regarding yield. To apply this data as labels for yield prediction, each different product type (feed, seed, and food) of a single crop was summed together and used as the actual total yield of that type for the given farmer in that year. The sum calculated for rye and rye wheat was then summed together, just as Engen et al. [12, 13], to keep the research efforts consistent. The data from 2012 until the season-ending in 2017 were used as supplementary information for yield models, representing historical yields.

4.4.2 Production Subsidies

Norwegian farmers and other agricultural enterprises rely on subsidies from the government as an income. The given subsidies are based on a farmer's production. Therefore, they are required to fill out an application each year concerning what type of agricultural activities they have performed and to what amount. These applications are used to create the production subsidies dataset, which contains information about the application for and payment of production subsidies for different agricultural enterprises in a given year. This dataset contains all subsidies applications in that given year, both paid out and applications currently being processed. The dataset consists of 174 features covering the area used for different purposes, the number of fruits and vegetables produced, the number of animals at the date of counting, and the amount and type of crops produced. As a farmer who grows organic crops is likely to receive more subsidies than a similar non-organic farmer, all 174 features are used to calculate the proper subsidies. The dataset also contains the calculated subsidy per category, such as subsidy for husbandry.

The noticeable features from production subsidies relevant for this research effort are those related to grain production. These are the total agricultural land area and how much of the land was used as arable land, surface-cultivated meadows, and infield pastureland, all listed in decares (daa). Along with these general land areas, the total cultivated land area for each of the grain types listed in the previous section was used to create a Kg of yield per decares (kg/daa) label. These production subsidies features are listed as the following in the datasets:

[9] https://data.norge.no/

- **p242** - Area used for barley
- **p243** - Area used for oat
- **p247** - Area used for autumn wheat
- **p240** - Area used for spring wheat
- **p238** - Area used for Rye and Rye Wheat
- **fulldyrket** - Arable land area
- **overflatedyrket** - Surface-cultivated meadows area
- **innmarksbeite** - Infield pastureland area

The features explained in this section directly describe parts of a farmer's grain production, such as the land area available. Because of the features information and the related theory in Chapter 2, these features likely strongly correlate to the final crop yield. On the downside, this information is not available before the harvest is completed, grains have been delivered, and the application has been processed. Therefore, they were used with care during the yield prediction experiments and might be more suitable as features representing historical production. Same as with the grain deliveries, the production subsidies were extracted from The National Data Catalog[10], first manually for exploration purposes and then automatically in our yield prediction system through their API.

[10] https://data.norge.no/

Chapter 5

Experiments

This chapter provides an overview of how our experiments were done, including data preprocessing and neural network design. Firstly we will describe how we managed to improve the results of Engen et al. [12, 13], which includes recreating their data preprocessing, training, and evaluation of their model. Further, the methods for crop prediction will be described, farm-scale, field-based, and a MIL model. Lastly, the recreated early yield prediction models with be presented.

The outline of this chapter is as follows:

- We begin with testing the first and second hypotheses in Section 5.1. This is done by including additional features and agricultural specific data that could lead to an increment of overall performance of grain yield prediction models.
- Further, we test hypothesis three to evaluate the usage of satellite images and vegetation indices for grain classification in Section 5.2. Additionally, we also experiment with hypothesis four to check whether multi-class attention-based deep multiple instance learning is a suitable method for semi-labeled data.
- Lastly, we test hypothesis five in section 5.3 to assess whether early yield predictions are successful without losing too much performance when using limited and in-season available data.

For each experiment, all the machine learning model architectures are explained. Additionally, training parameters and settings will be presented. The results of each experiment will be described in chapter 6.

5.1 Baseline Approach: Improving on Grain Yield Predictions

Our baseline approach revolves around extending the training data from the previous student by expanding the dataset with features and data samples. The following sections describe how different methods were tested separately by extending their previous scripts.

5.1.1 Implementing ANN for Additional Features

After obtaining access to the recent work done by Engen et al. [12, 13], our first task was to rerun their code and getting everything set up. To start testing our first hypotheses, their code-base was extended to include the three new features in the training dataset: *sunlight duration*, *state of the*

ground, and *growth degree*. We merged these features on organization number and year, similar to the previous work, which extended the number of features from 900 to about 1500. To further address our hypotheses, a metric measuring the soil quality of each field for each farm were included. The following features were added on a per farm basis:

1. A mean value of the soil quality for all fields
2. The number of fields with a soil quality of 1
3. The number of fields with a soil quality of 2
4. The number of fields with a soil quality of 3

The creation of this data and its structure were described in Section 4.2.3. After extending the dataset with the additional features, the simple dense model from Engen et al. [12, 13] were trained. The model tended to overfit during the first training process, which resulted in some minor model adjustments. The first dense layer in the model was halved in size to 256. Lastly, the dropout percentage in the dropout layer was slightly increased to 0.25 and 0.5, respectively. This refined model is used in all experiments where the dataset was extended only with numerical data.

5.1.2 Implementing ANN for Additional Data Samples

Our second hypothesis states that adding more data samples to the training dataset would increase accuracy. To evaluate this hypothesis, new samples would need to be acquired and applied. To achieve this, the Frost API scripts were extended to include data from 2020. The new features were run through the same scripts for cleaning, interpolating, and normalizing, which guaranteed that the new data was structured identically to the old. This dataset with new samples were also applied to the simple dense model, similar to the previous experiment.

5.1.3 Implementing ANN for the Combination of Additional Features and Data Samples

To further explore the hypotheses mentioned above, another experiment were conducted which combined the two previous ones. The methodology behind this experiment is mostly the same as the previous ones. We added more samples and more features at the same time, and used it to re-train the model in the exact same way.

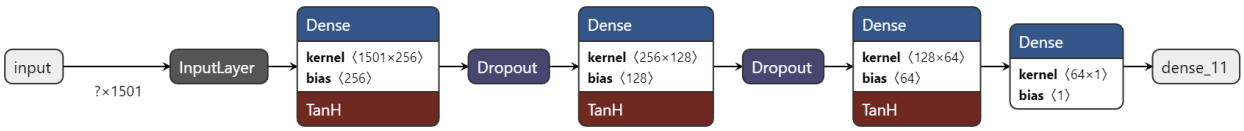

Figure 5.1: Refined dense model visualization

Figure 5.1 shows the refined dense model used in all three experiments explained above. The figure shows that the model only contained four dense layers using tanh as an activation function except the last one. Dropout layers were added after the first and second dense layers. The input layer was change to match the size of the training dataset. An early stopping callback tracking validation loss is also added with patience of 100, which means the model stops training if it does not improve validation loss for 100 epochs. Adam was used as the optimizer with a learning rate of 0.001. This architecture of the model, along with all other to come, were visualized using Netron[1].

[1] https://www.electronjs.org/apps/netron

5.1.4 Implementing Hybrid Model for Additional Features

Lastly, we ran one final experiment using the best model from the previous work, the hybrid model, which uses satellite images. As explained in Chapter 4, we could not obtain satellite images for 2020 because of financial limitations. Therefore, we recreated the hybrid model by only adding more features. Our numerical dataset contained 1500 features similar to our first experiment and was merged with the satellite images during training. Each satellite image series was coupled with a yield value, weather features, and other custom features. The images were also augmented with salt and pepper, stride-augmentations, and random rotations. All of these functions were applied during training using a TensorFlow Generator [2]. The model was trained for 100 epochs using the structure shown in Figure 5.2

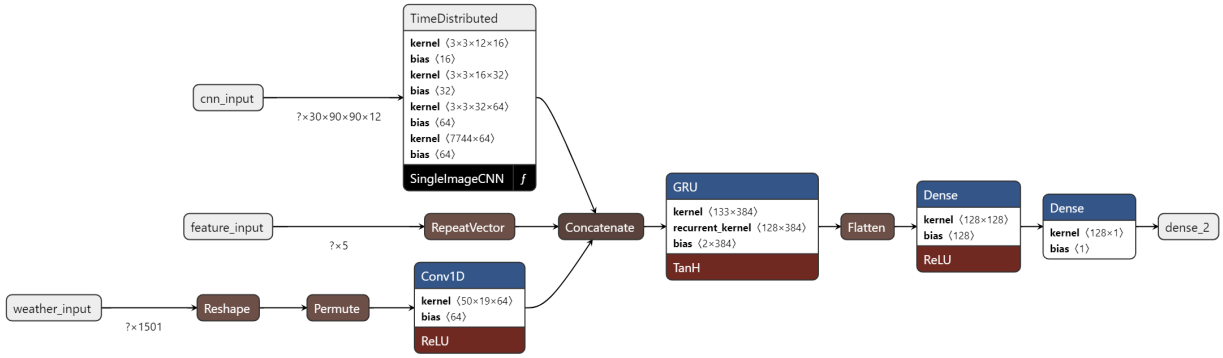

Figure 5.2: Hybrid yield model visualization

As shown in Figure 5.2 each of the features were inputted separately before being concatenated and fed into the GRU layer. Lastly, the data is flattened and fed into two dense layers. Adam was also used as the optimizer but with a learning rate of 0.001. This model also implemented an early stopping callback with a patience of 10. The results obtained from all of the experiments mentioned above will be described and modeled in Chapter 6.

Our baseline method of extending the data applied to grain yield prediction improved the predictions. However, this does not change the fact that these grain yield prediction techniques cannot be implemented in Norway's agriculture. This is because the techniques apply data from the whole growing season, meaning that the prediction model cannot be used before the growing season is over. While studies like Engen et al. [12, 13] have attempted early yield prediction, this was without knowing the type of crop planted on each field. In order to apply the techniques which were improved in this section, crop classification needs to occur, which is what we will implement in the forthcoming section.

5.2 Novel Approach to Grain Classification Using Spectral Temporal Features

This section describes how we performed grain classification using spectral and temporal features. The data applied were a set of multi-spectral band images representing a specific geographic location in a given time. These images were either the raw satellite images or the derived vegetation indices, as explained in Section 4.1, Chapter 4. Firstly a farm-scale grain classification was implemented, followed by a field-based grain classification. Temporal profiles for each of the ten vegetation indices will be performed to evaluate the indices and identify optimal periods for classification. This is followed by field-based classifications using vegetation indices. Lastly, a multiple instance learning

[2] https://www.tensorflow.org/api_docs/python/tf/data/Dataset#from_generator

model for multi-class semi-labeled data will be implemented.

5.2.1 Implementing Hybrid Model for Farm-Scale Grain Classification

The farm-scale grain classification is similar to the yield prediction using the hybrid model and satellite images. The only difference is that only satellite images are used, not any numeric data. The satellite images used for this experiment were described in Section 4.1.1. Firstly the images are split into training and validation sets using a split ratio of 0.8, and all images are labeled using the dataset described above. Further, the images are loaded as a TensorFlow Generator, which only loads images in batches into memory during training. This means that, the masks and other augmentation techniques are applied during training. Our initial farm-scale crop classification model is shown in Figure 5.3.

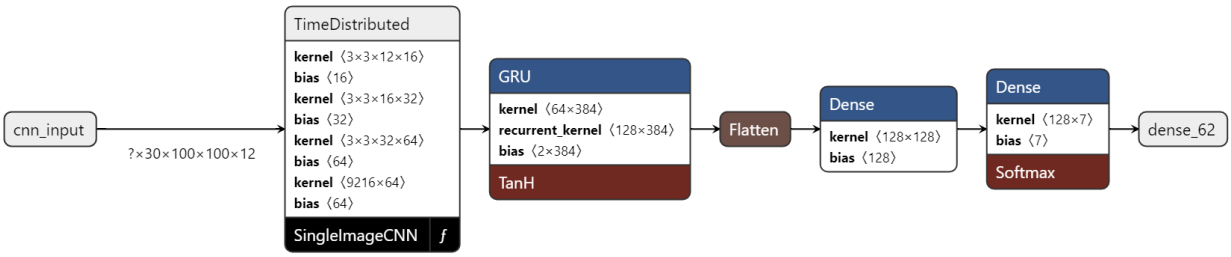

Figure 5.3: Farm-based crop classification model visualization

Figure 5.3 shows that only the time-distributed part of the hybrid baseline model is used. The vector from the time-distributed layer is fed into a GRU encoder which encodes the entire sequence into a 64 length vector. Lastly, the vector is flattened, and the last dense layer of the model using a softmax activation function yields a vector with percentage predictions for each crop type. However, this model did not perform as well as expected. Therefore, some measures improvement had to be implemented

To improve this model, which struggled with overfitting, hyperparameters were tuned, classes with low occurrences were removed, and the number of images to use as time series were limited. In addition exponential learning rate decay using different parameters to prevent further overfitting were also enforced. These adjustments did not produce any improvements, which resulted in an additional reason for performing field-based crop classification.

5.2.2 Implementing Optimized Hybrid Model for Field-Based Grain Classification

The need for field-based image classification have been presented throughout this research effort, and were strengthened by the low performance of the farm-scale crop classification. Even if the field-based classification were to achieve high performance, one could not simply apply an image of a random farm to this model. This is because farmers generally does not produce the same crop on all of their fields. Which is why a machine learning model needs to learn the distribution between fields and crop types, and not between farms and crop types.

Chapter 4 explained the creation of field-based satellite images. The number of images used in a field-based grain classification model is significantly higher than that of a farm-based one. The field-based images were of size 16 x 16 x 12, which is big enough to contain even the largest fields, while still being significantly smaller than the original size.

Before training, the images were loaded and coupled to their belonging label. We opted to remove

fields with a small area because our smallest fields would only cover one or two pixels in the satellite images, which is a tiny amount of data to feed the network. This was done for all future experiments concerning field-based crop classification. Further, the new satellite data were augmented using salt and pepper, rotation, and random flips. Augmentation was implemented due to its effectiveness in counteracting overfitting in machine learning. Lastly, the images were inserted into a TensorFlow Generator and fed into the model. In this experiment, while re-creating the model for field-based images, Ferlet [101] was used as inspiration to improve the model further. In addition, Kussul et al. [14] successful 2D CNN implementation with multiple sets of CNN, max-pooling, and ReLu were evaluated. As explained in Chapter 3, Kussul et al. [14] were able to classify crop types with an accuracy of 94.6% successfully. Their choice of architectural structure and functions were therefore evaluated for our setting.

This model had too many layers and neurons, so some were removed and the remaining layers were made smaller to neutralize overfitting. The final model is described in Figure 5.4

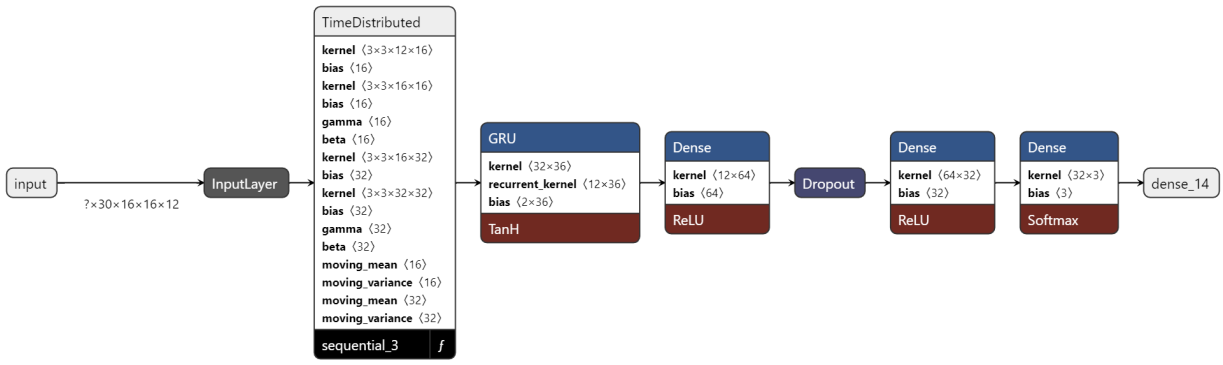

Figure 5.4: Field-based classification model

As shown in Figure 5.4, this model is similar to the farm-based model. Some differences are that the time distributed part has more parameters like moving mean and variance, beta, and gamma, inherited from the batch normalization layers in the CNN net. Further, an extra dense layer is added after the GRU layer, including a dropout layer. These dense layers are also smaller than the farm-based model's dense layers. Lastly, the most significant difference is the GRU layer's size of 12 units instead of 64, which impacted the model when preventing overfitting.

5.2.3 Implementing Hybrid and Optimized Hybrid Model for Grain Classification using Vegetation Indices

By the time the vegetation indices had been developed, farm-based crop classification in Section 5.2.1 had already been performed and the scrip for creating small images was completed. Considering the bad results, farm-based classification using vegetation indices where excluded from this research effort. This choice was also made considering the models limited generalization abilities as farmers often grows different types of crops in a growing season. Field-based images of size [30 x 16 x 16 x 10] were created, as explained in Chapter 4, along with a corresponding field mask. These images where then applied to the initial classification model from Section 5.2.1 for two experiments. One where all seven crop types were the target classes, and one where only the five crop types relevant for yield prediction where used. Both experiments only used the vegetation indices derived from the satellite images between week 15 and 30. The earliest images where excluded because many of the early satellite images where found to contain a lot of snow, cloud, atmospheric disturbance and other type of noise, and the latest images where excluded as the final goal of these experiments where to identify crop type early in the growing season.

At this time, the optimized classification model from earlier in this chapter had been created. To

ensure the whole data distribution is utilized, all indices derived from the five crop types every week were applied to the optimized model to present as many features as possible. With still no improvements, the findings from Peña-Barragán et al. [82] presented in Chapter 2 were adopted. They found late-summer and mid-spring images to contribute to 60% and 30% of the classification learning, respectively. Our first two experiments on crop classification using vegetation indices used images between weeks 15-30, which covered most of the mid-spring and early-summer period. To therefore account for the 60% contribution of the latest weeks, a fourth experiment was performed using images from weeks 30 to 39.

Exploring Vegetation Indices

The validation loss and accuracy rarely changed while training the model during the first four experiments with vegetation indices. Even optimizing the model, which was successful for the classification of raw images, did not improve the performance. Therefore, more optimal features needed to be identified. Inspired by the works of Foerster et al. [16] on plotting and studying vegetation indices throughout the growing season, as explained in Section 3.4, temporal profiles for each vegetation indices were created. This was done by calculating the average values for each of the ten indices in every image. This was done through the following steps:

1. The farmer's field masks from Section 4.1.2 in Chapter 4 were applied to their respective locations to only use the vegetation indices values from fields.
2. The values from one vegetation index image were normalized between *0* and *1*, and the average was calculated and saved. This was done for every vegetation index image, meaning it was done for each of the farmer's 30 images every year.
3. The averages calculated in the previous step were collected across all images to calculate an average of the averages. This was done while filtering on crop types, meaning that each vegetation index ended up with an average value for each week for each crop type.

These steps resulted in an average vegetation index value for each crop type in each satellite image. These temporal profiles were plotted for analytical purposes, which enables us to understand the development of the different indices throughout the growing season. This enabled us to evaluate each vegetation indices potential for crop classification. Based on this knowledge, the indices were further explored to identify the most optimal period and that period's most distinguishable vegetation indices. The vegetation indices will be shown in Chapter 6.

These identified optimal vegetation indices and periods were extracted and applied to the initial and optimized classification model for evaluation purposes. Surprisingly, these experiments resulted in the same accuracy as the previous ones. This was likely because of overfitting, just as identified for the raw satellite images.

During each experiment on crop classification using vegetation indices, a range of different learning rates, layer sizes, and crop and time intervals were tested, both with and without masks, with no noticeable changes. The validation loss and accuracy from each significant experiment will be presented in Chapter 6.

5.2.4 Implementing Multi-class Attention-based Deep Multiple Instance Learning

Hypothesis 4 states that a Multi-Class Attention-Based Deep Multiple Instance Learning (MIL) model is a more suitable method of crop classification for semi-labeled data. This technique is therefore implemented in order to evaluate our hypothesis and to reliably use our semi-labeled dataset.

This could enables us to more accurately predict crop type on all our field data instead of relying on our field-based crop classification model. The Attention-Based MIL model were implemented using Keras' implementation using TensorFlow and Keras[3]. One downside of this implementation is that we could not get our TensorFlow Generators to work in the network. Therefore only a portion of our satellite images were directly loaded into memory before training. Further, the images were split into training bags and validation bags using different parameters and a bag size of 7. The bag size of 7 was selected because this created the best balance between positive and negative bags. The network were trained with the Adam optimizer using 0.001 as the learning rate and 150 epochs. The bag sizes tried were 100, 200, 300, 500, and 750. Additionally, validation bag sizes were varied between 100 and 250 depending on how the training bag size.

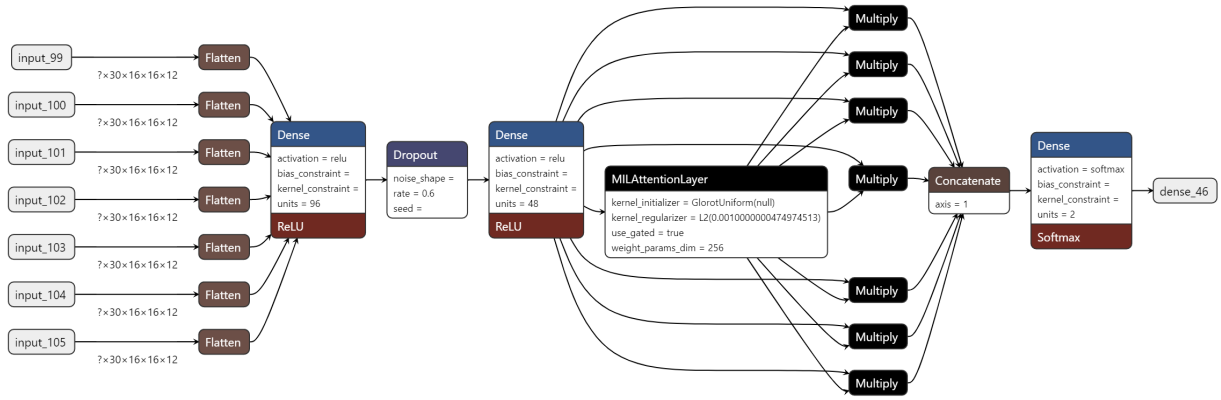

Figure 5.5: Visualization of the attention based deep multiple instance learning model

Figure 5.5 shows the model architecture of the MIL model implemented. Each input layer is assigned to each bag fed into the model. Each bag contains an image series of 30 images of fields, which are flattened before being forwarded to two dense layers with a dropout layer in between with a rate of 0.6. The next step is the attention layer which calculates the attention scores before concatenating. Lastly, the data is inserted into the last dense layers of two units using a softmax activation function. Two units are used because each bag is classified as either negative or positive.

Given that this MIL network only classifies a bag as positive or negative, the system was extended for the purpose of multi-class. One model was created for each crop type, where the difference between the models was which class was marked as positive. Each model was trained multiple times using different bag settings. To correctly classify fields in the system a field would have to be predicted through each model. The field are then given the label of the model where the field have the highest attention scores in a positively predicted bag. The fields with the highest attention scores are the fields that contributed to the prediction. Therefore, these fields are classified as the value of the positive label used for that specific model.

After improving the existing yield prediction methods and successfully classifying crop types, these techniques can be linked together to perform early in-season yield predictions. The hypothesis concerning early yield prediction states that early yield should be achievable only using data available in season, which means that crop prediction, used to create features for early yield prediction, was a prerequisite. After conducting several experiments to achieve satisfactory results, we conclude this part of our research effort with the MIL classification model.

[3] https://keras.io/examples/vision/attention_mil_classification/

5.3 Novel Approach to Early Grain Yield Predictions

5.3.1 Implementing Early Crop Type Classification

For the early prediction experiment, we aim only to use satellite images from weeks 10 to 21 and 10 to 26. This type of experiment creates a model which can be used to predict classes on labeled and unlabeled data, and further, the model's output will be used as input data to our recreated yield prediction models. The problem is that is it not able to train on unlabeled data, which our MIL model achieves. The unlabeled data is the fields that we did not include in the classification data described in Section 4.2.4. Another reason for conducting this experiment is that creating a model which uses data acquired early in the growing season would be more applicable to real-life scenarios. In comparison, a model that requires data from week 30 of the growing season would not be of much use. It also removes the need to fetch the crop type from data that would not be available until long after the growing season.

The data pre-processing, model implementation and training are identical to the field-based crop classification described in Section 5.2.2. The only difference is the number of images fed to the network. Hence the network's input shape also changes accordingly. The reason for not using the MIL model to predict crop type is that it did not produce concrete results easily usable for the early yield models. This is why we opted to use the field-based classification model as template for our early crop type prediction model. Inputting different amounts of satellite imagery to the network is done by changing the iterator supplied to the TensorFlow Generator. Instead of yielding the entire time series of 30 images, only images from week 10 to 21 or week 10 to 26 are returned. Both classification models were trained using the exact same parameters as our field-based grain classification model.

The trained model created from this experiment was then used to predict crop types on all our field data, excluding those having a small area. After our trained model successfully predicted crop type on all fields, we created a file containing information about what field planted which crop type. This file would further enable us to create masks for each farm for each crop type so that the yield prediction model only utilizes the correct fields when predicting yield for a specific crop type. It would also serve as a lookup table for the crop type input feature that we will use in our novel approach to early yield prediction, described in the next sections.

5.3.2 Implementing Early Grain Yield Prediction

By performing this experiment, our goal is to recreate the previous work done by Engen et al. [12, 13] where they attempt early predictions. We used their experiment as a template but included our new features in the same way as presented in Section 5.1.1. We also added the predicted crop type as a feature by using the organization number, year, and field as an identifier. Two models were created similar to our early crop type prediction models. The reason behind this is to explore hypothesis 5 which states that in-season early predictions are just as successful when using predicted crop types. Crop types need to be predicted because crop type data is not available during the growing season, due to it being extracted from delivery data. The entire dataset, including satellite and numerical data, contains data for a whole growing season, ergo all 30 weeks. We want to limit the dataset to include data from week 10 to 21 and week 10 to 26. Next section will describe how this filtering is done

Given that some numerical data has multiple features each day, we must include data from day one of the growing season to day 84 (12 * 7). This data removal is done by filtering the dataframe containing the numerical data by column names using a regular expression. Each feature containing daily data was filtered while combining it with the satellite images, as shown below in Figure 5.6. This example shows how we only included data from columns named after the feature name

combined with day number from 0 to 84. This corresponds to week 10 to 21. The regular expression used for data from week 10 to 26 was $min_temp(([0-9])|([1-9][0-9])|(10[0-9])|(11[0-8]))$, which filters the data from day 0 to 119. The text min_temp was swapped out to match each of the weather features identical to Figure 5.6

```
min_temps = wd.filter(regex='min_temp(([1-7]?[0-9])|(8[0-3]))$').values
mean_temps = wd.filter(regex='mean_temp(([1-7]?[0-9])|(8[0-3]))$').values
max_temps = wd.filter(regex='max_temp(([1-7]?[0-9])|(8[0-3]))$').values
total_rain = wd.filter(regex='total_rain(([1-7]?[0-9])|(8[0-3]))$').values

sunlight = wd.filter(regex='sunlight(([1-7]?[0-9])|(8[0-3]))$').values
daydegree5 = wd.filter(regex='daydegree5(([1-7]?[0-9])|(8[0-3]))$').values
ground = wd.filter(regex='ground(([1-7]?[0-9])|(8[0-3]))$').values
```

Figure 5.6: Example usage of filtering weather features using regular expressions.

After all data preprocessing was done, we trained the model similar to the previous students, which is their hybrid yield prediction model described in Figure 5.2. The only differences are the model's input shapes.

Chapter 6

Results

This chapter presents the results achieved from the experiments conducted in the previous chapter. Section 6.1 explains the improved grain yield prediction models and how different data additions affected the model. Further, section 6.2 presents the results from our novel approach to grain classification. The section includes results from farm-based, field-based, vegetation indices, and MIL models. Lastly, section 6.3 presents how the early yield model from Engen et al. [1] was reproduced using fewer features. Our novel approach to early yield prediction achieves comparable results to the initial model but is way more applicable to real-life scenarios of in-season predictions.

6.1 Baseline Approach: Improving on Grain Yield Predictions

This section covers the results from the baseline experiments. Multiple methods for improving the results were implemented, including adding more data samples, adding more features, and adding both more data samples and features simultaneously. The model reimplemented was the simple dense model from Engen et al. [12, 13], which is described in their research effort as Weather DNN. Their results were reproduced to evaluate our reimplementation and compare their results with ours. Since as satellite images for 2020 were not acquired, the only way to improve their hybrid model was to include more features. Results from all of these experiments will be presented below.

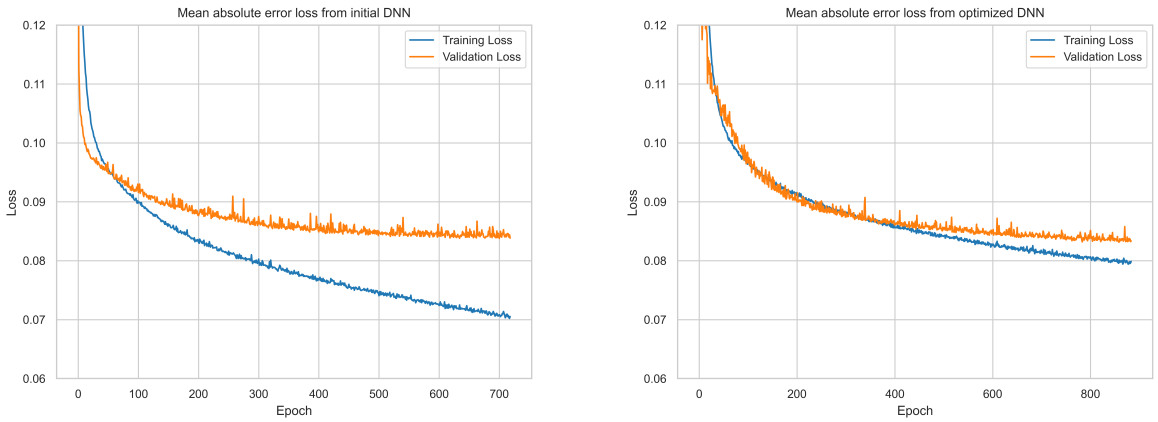

Figure 6.1: Graphs showing loss achieved from the original DNN model vs the optimized DNN model

The left graph in Figure 6.1 shows training loss and validation loss for the original model reimplementation. As shown in the figure, the model tends to overfit. To prevent overfitting, the first dense layer in the original model was halved in size to 256, and the dropout percentage in the dropout

layer was slightly increased to 0.25 and 0.5, respectively. The results from this implementation can be seen in the right graph, where the gap between training loss and validation loss is significantly smaller, proving that overfitting is reduced. The Mean absolute error measured in Kg/daa achieved by the original and optimized models was 83.85 Kg/daa and 83.28 Kg/daa, respectively. For further experiments, the optimized model was used.

6.1.1 Results on Additional Features Applied to ANN

After the three extra weather features were added as described in Chapter 5, the number of features was increased to 1500. The optimized model using an increased input shape was trained for 1200 epochs. The loss graph is shown in Figure 6.2

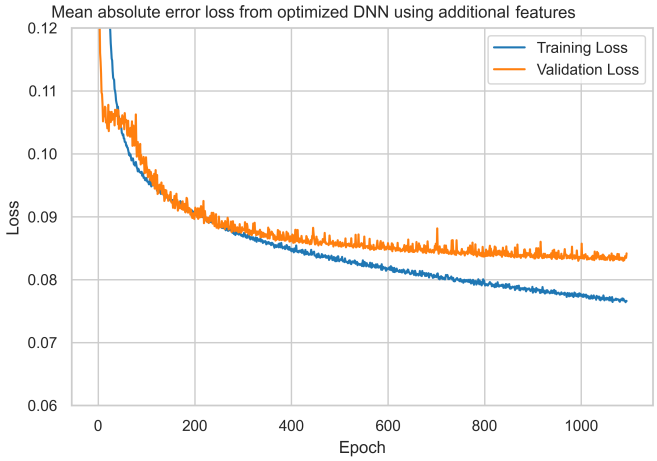

Figure 6.2: Yield prediction loss achieved from the optimized model when adding more features

Figure 6.2 shows similar loss values compared to Figure 6.1, with a slight decrease, although this is hard to notice from the figures. The MAE was calculated to 82.12 kg/daa, an decrease of 1.16 kg/daa. The number of epochs used during training was increased to 1500 because the early stopping callback did not trigger using the increased training dataset.

6.1.2 Results on Additional Data Samples Applied to ANN

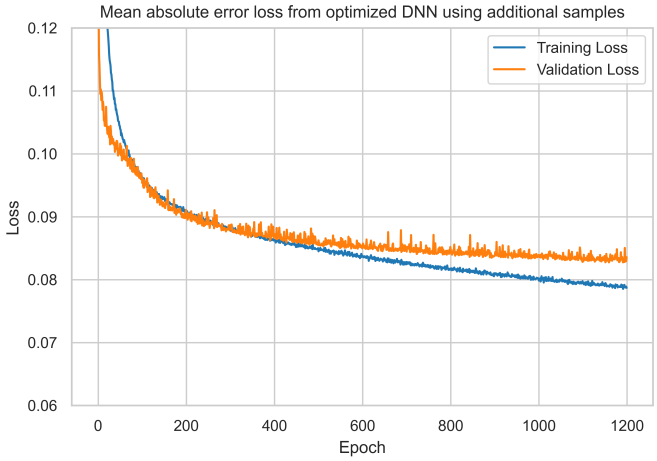

Figure 6.3: Yield prediction loss achieved from the optimized model when adding more samples

After extending the training data as described in Chapter 5, the training data size increased by 33%, and the optimized model was trained for 1200 epochs achieving the results shown in Figure 6.3. The figure shows the training loss and validation loss achieved from the optimized DNN model using data including 2020. As we can see, both losses decrease. We also calculated the MAE for this experiment to 82.82 kg/daa, almost identical to the optimized model.

6.1.3 Results on Additional Features and Data Samples Applied to ANN

After adding more data samples and more features, as described in Chapter 5, the dataset had 600 more features and approximately 15 000 additional samples. The optimized model using an increased input shape was trained for 1200 epochs. The loss graph for this experiment is shown in Figure 6.4.

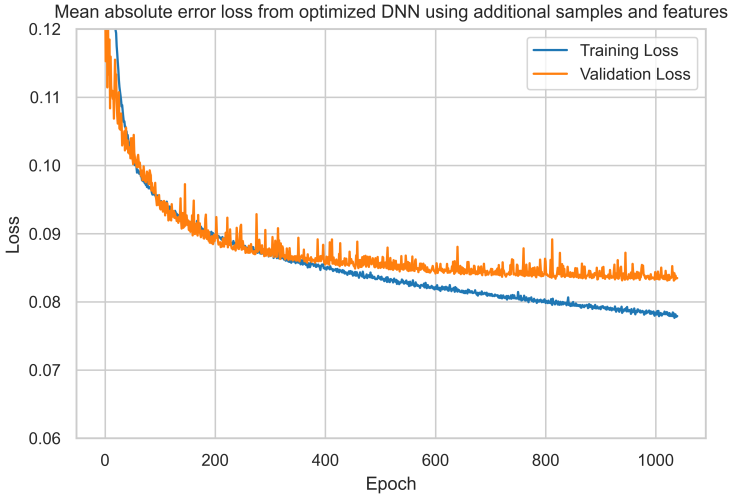

Figure 6.4: Yield prediction loss achieved from the optimized model when adding more samples and features

Figure 6.4 shows the optimized model's training and validation losses when adding more samples and features to the training dataset. The loss lines have a similar trend as the previous loss graphs, showing a decrease in loss compared to Figure 6.1. The MAE was calculated to 82.60 kg/daa, which shows a slight decrease of about 0.7 kg/daa. The time it took to train each of these models was slightly increased when adding more samples and features. However, training this simple DNN model took about 5 to 6 minutes. Thus total training time was low.

6.1.4 Results on Additional Features Applied to Hybrid Model

The final baseline experiment on improving yield predictions was to apply the new features to the previously best-performing hybrid model. Figure 6.5 shows the hybrid model's training and validation loss when applying the new features. Keeping in mind that this experiment only ran for 70 epochs, the same trend of the loss graphs can be identified in our previous experiment. The MAE was calculated to be 78.9 kg/daa, which is a increase of about 2.53 kg/daa compared to the best results achieved on the hybrid model by Engen et al. [12, 13]. The MAE of the hybrid model with new features was calculated using its latest iteration with a loss of 0.079 and not the best iteration, which had a loss of 0.076. This is simply because the best iteration was never saved. Training the hybrid model took about 80 consecutive hours, which is why the experiment was not redone for the exact MAE.

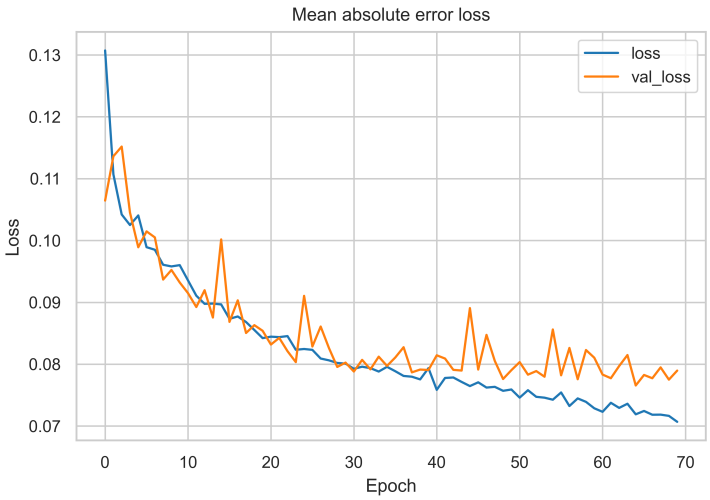

Figure 6.5: Graph showing training and validation loss achieved from the hybrid model when adding new features

Experiment	Validation Loss	MAE kg/daa
Initial Model	0.90	83.85
Optimized Model	0.87	83.28
Additional Features	0.82	82.12
Additional Samples	0.84	82.82
Additional Samples and Features	0.83	82.60
Hybrid with Additional Features	0.79	78.96

Table 6.1: Validation loss and MAE achieved for each of the yield prediction models implemented.

Table 6.1 summarizes the results for every baseline experiments on improving the yield predictions, for the purpose of comparison. The experiments are shown in the order in which they were conducted.

6.2 Novel Approach to Grain Classification Using Spectral Temporal Features

6.2.1 Results on Farm-Scale Grain Classification using Hybrid Model

Our farm-scale crop classification model aims to classify what was planted on a given farm based on the time series of 30 images. The classification dataset assumes that all the fields connected to a farm contain a particular crop type. This assumption could lead to questionable results because the classification dataset was not created based on ground truth, as described in chapter 5. The images could also contain a considerable amount of noise, which could mislead the classification model. However, we used the CNN and GRU part of the hybrid model described in chapter 5 for our first classification model.

Since the farm-scale experiment results were unsatisfactory, we tried different approaches to improve the model. Firstly we tried inputting one image at a time from the time series of 30 images. Further, we tried including all 30 images with seven and three classes, respectively. The reason for not

including all seven classes is the imbalance of the labels, as seen in Figure 6.6. Lastly, we tried adjusting the network parameters to make them smaller.

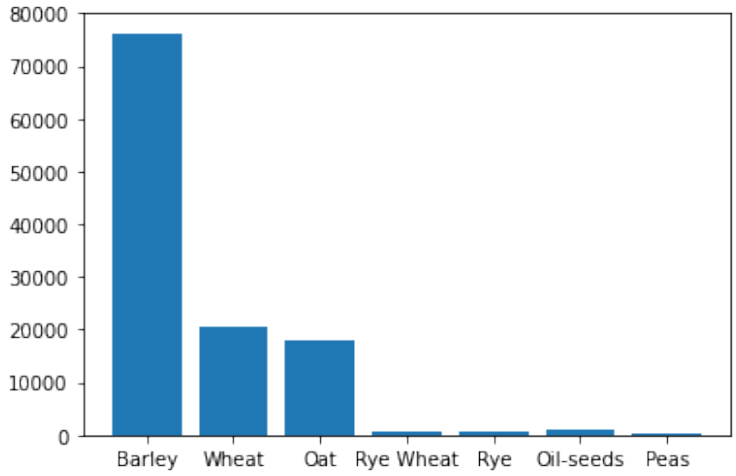

Figure 6.6: Bar chart showing the class distribution of the classification dataset

Figure 6.6 shows the distribution of classes in the classification dataset. As we can see, barley, wheat, and oat is heavily overrepresented. Removing the four least represented classes would make the dataset more balanced and could have improved the model.

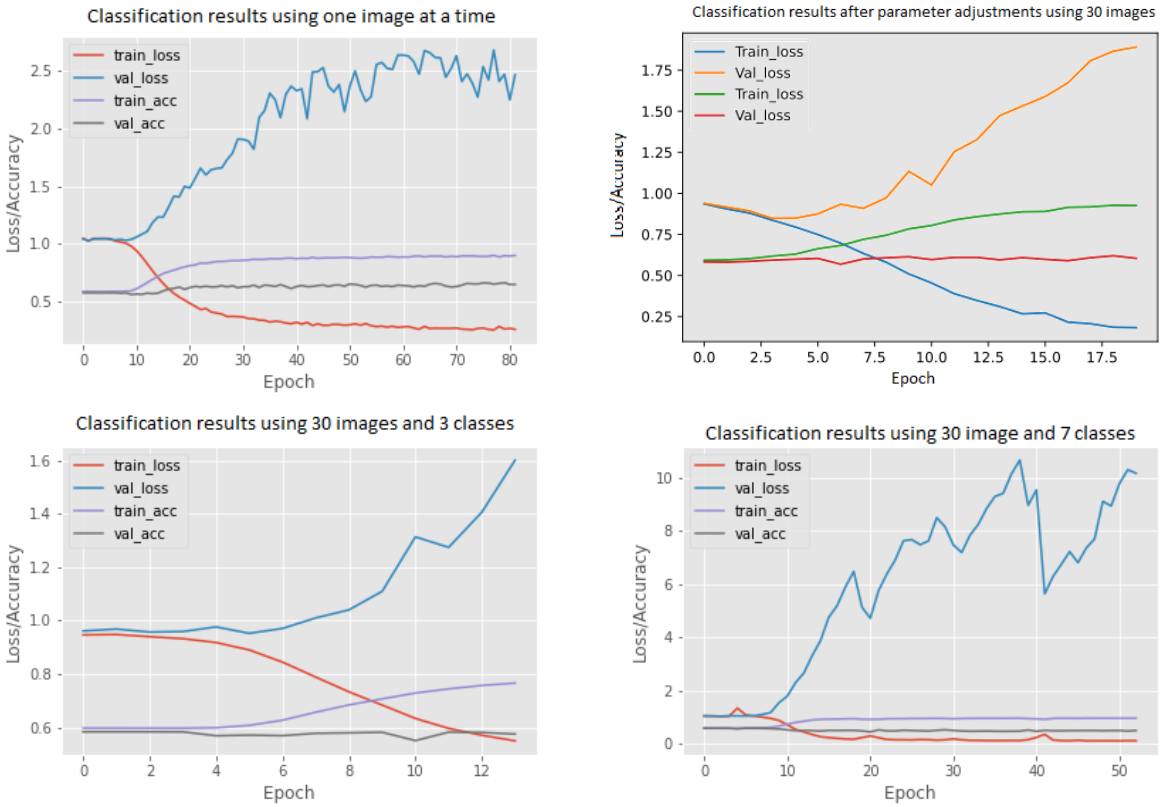

Figure 6.7: Graphs showing results from different approaches on farm-scale classification

Figure 6.7 shows four different experiments on farm-scale crop classification. The training loss generally has a downwards trend throughout the experiments, meaning that the model manages to learn some features. This also implied that the training accuracy increased. Furthermore, the validation accuracy primarily stays constant, even after training for 80 epochs, and the validation loss increased in most cases. The different models were trained for a different amount of time because each experiment started showing the same negative trend as the first one, which was the

model top left in Figure 6.7. The graphs show that the validation accuracy resides around 0.6 for all of our implemented models. The highest training accuracy achieved was 0.8 by the random image model, though this model overfits.

6.2.2 Results on Field-Based Grain Classification using Optimized Hybrid Model

The field-based crop classification model aims to classify crop type based on time series images of one field instead of the entire farm. The advantage of using field-based images instead of farm-based is that the images contain less noise than the original images. Thus, the training process will be faster. The creation of the field images is described in chapter 4. Our initial model copies the farm-scale crop classification model with different input parameters. The results achieved by this model are almost identical to the farm-scale classification model, which was a training accuracy of 70% and a validation accuracy of about 60%.

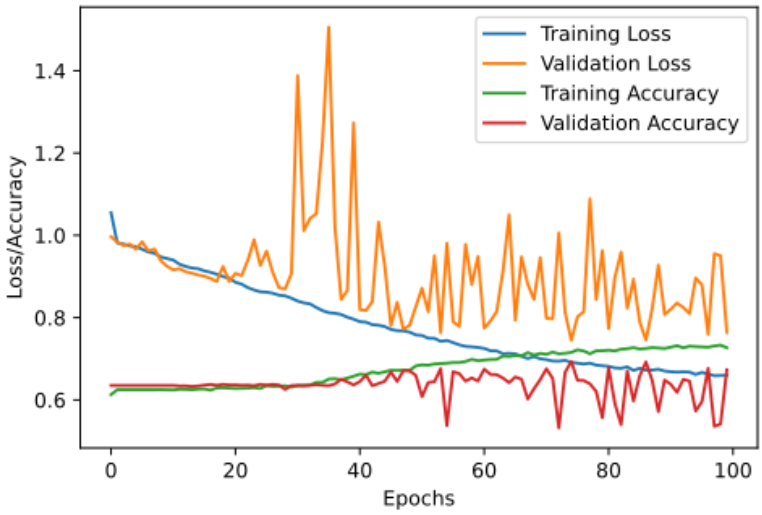

Figure 6.8: Graph showing results from crop classification using 16 x 16 images

Figure 6.7 shows the results from training our optimized field-based crop classification model. As seen, both losses decrease, and both accuracies increase, which is already an improvement compared to the farm-scale model. At best, this model achieved a validation accuracy of 70%, an improvement of 10%. These results prove that the smaller images containing less noise serve better for crop classification. We can also see that the validation loss and accuracy fluctuate after 20 to 30 epochs. The reason behind this will be discussed in the next chapter. Most of the results from crop classification show tendencies of overfitting which were identified multiple time as a result of our data distribution. Especially when the datasets included satellite images.

6.2.3 Results on Field-Based Grain Classification using Vegetation Indices

Results on Analyzing Vegetation Indices

As explained in Section 5.2.3, the average vegetation indices throughout the growing season were calculated for each crop type. With so many crop types plotted on ten different graphs, one for each vegetation index, the graph can quickly get cumbersome to read and understand. Oil seeds and pea were excluded from the results to remove less relevant information as they are not relevant for either crop classification or crop yield. Figure 6.9, therefore, shows the graph for only the five main crop types. Each graph in the figure is the temporal profile of the specified vegetation index, where

each plotted line represents a crop type. Some of these graphs might seem identical at first, but when explored in detail, we extracted meaningful information about our vegetation indices. Many of the indices showed a similar trend in the temporal development, where they would slowly grow during the growing season up until a point. This point is likely to be close to when the crops reach maturity, and then they rapidly decline during what is likely to be the period of harvest.

While the development and curve of the vegetation indices would be interesting to study in terms of grain yields, it is the difference between each crop type in a given week that is interesting for crop classification. By studying each graph and the difference between crop types, a trend can be seen between weeks 18 and 25. In this period, the indices tend to differentiate between each crop type. Through feature elimination, time periods and vegetation indices were eliminated. This process was done until an optimal combination was identified where each index had a clear differentiation between each crop type over the whole period. These indices in the optimal time period are shown in Figure 6.10. The figure shows the SIWSI, LSWI, NDRE, and GRNDVI between weeks 19 and 24. Not only are the difference between each crop type both significant and consistent, but the graphs also rarely cross each other paths.

The temporal profiles of all crop types, including oil seeds and peas, can be found in Appendix A, but it should be noted that oil seeds and peas only had 16 and 3 locations, respectively, to calculate the average.

Results on Field-Based Grain Classification using Vegetation Indices on Hybrid Model

Chapter 5 explained six different experiments on applying vegetation indices to crop classification. The results from these experiments are explained in Table 6.2. The two first experiments are on the initial model, excluding the first and last images. This is where the lowest loss and highest accuracy were identified, but these are only minor improvements from the other experiments. The figure presents the setting of the different experiments. The five crop types are the main crop types applied for yield prediction, while all types represent these five main types, along with oil seeds and peas. The week interval represents which weeks the applied images are from and the number of images, having one image each week.

Nr	Features	Model	Crop Types	Week Interval	Val Loss	Val Accuracy
1.	All	Initial	Five Types	15-30	**0.8910**	**0.6423**
2.	All	Initial	All Types	15-30	0.9745	0.6330
3.	All	Optimized	Five Types	10-39	0.9222	0.6366
4.	All	Optimized	Five Types	30-39	0.9247	0.6353
5.	Optimal	Initial	Five Types	19-24	0.9498	0.6359
6.	Optimal	Optimized	Five Types	19-24	0.9520	0.6353

Table 6.2: Crop classification experiments using vegetation indices

The results of the third experiment are the attempt to apply all vegetation indices for the whole growing season. At the same time, the fourth is the experiment using the last ten images in the growing season. Lastly, the results from the fifth and sixth experiments in Table 6.2 are from the model trained using the optimal vegetation indices and time periods identified in the previous section. Applying the optimal vegetation indices resulted in what we can consider no change in the validation loss and accuracy. Surprisingly, our efforts in improving the classification model and optimizing the applied data distribution for vegetation indices resulted in zero improvements. This is likely because of factors such as noisy satellite images and limitations in the vegetation indices. These factors will be thoroughly discussed in the next chapter. As mentioned in Chapter 5, a wide range of additional experiments was performed between the ones presented here, where small changes were made to the input data and hyperparameters. These small experiments also

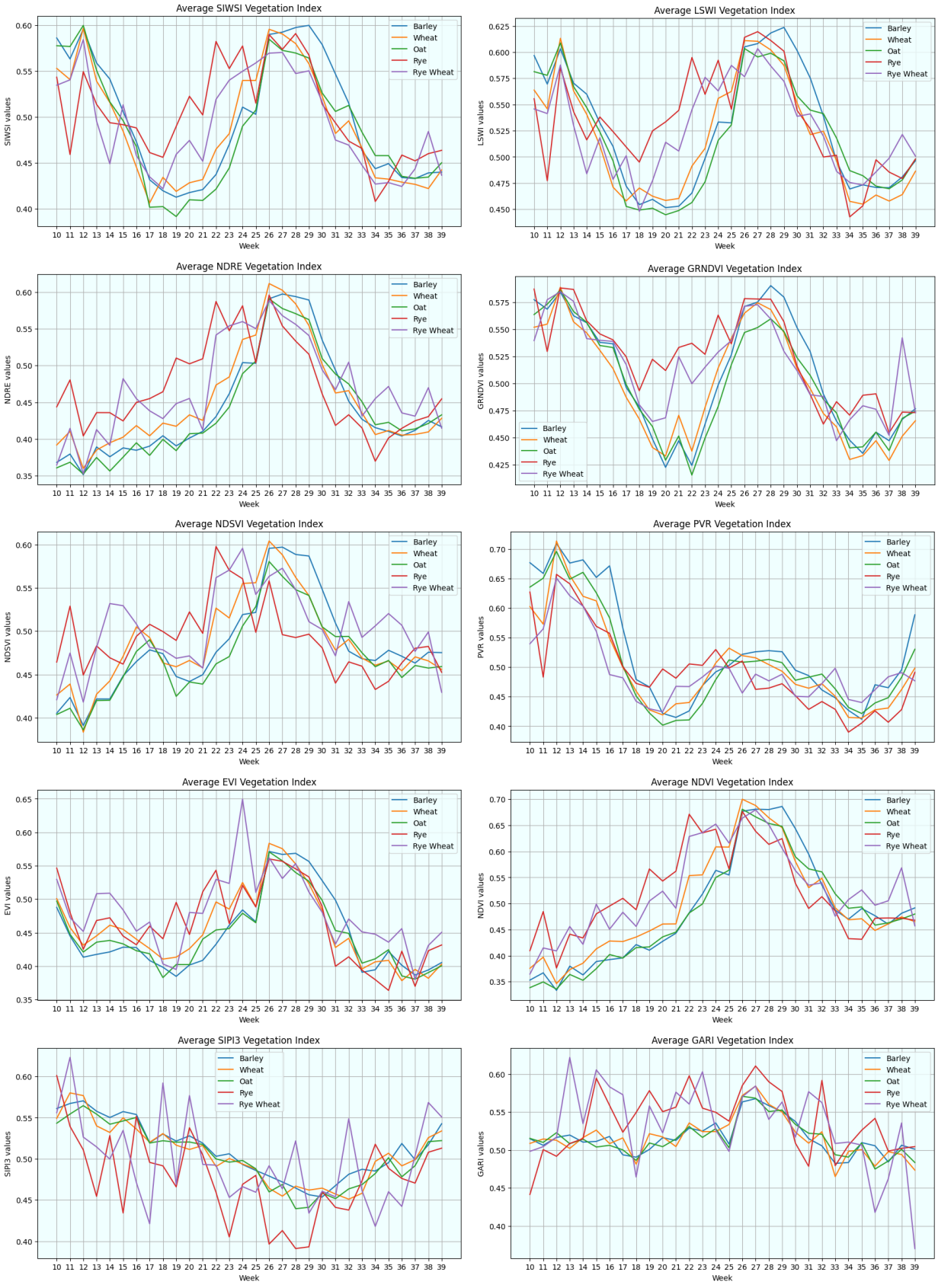

Figure 6.9: Average vegetation indices for the five relevant crop types, throughout the growing season

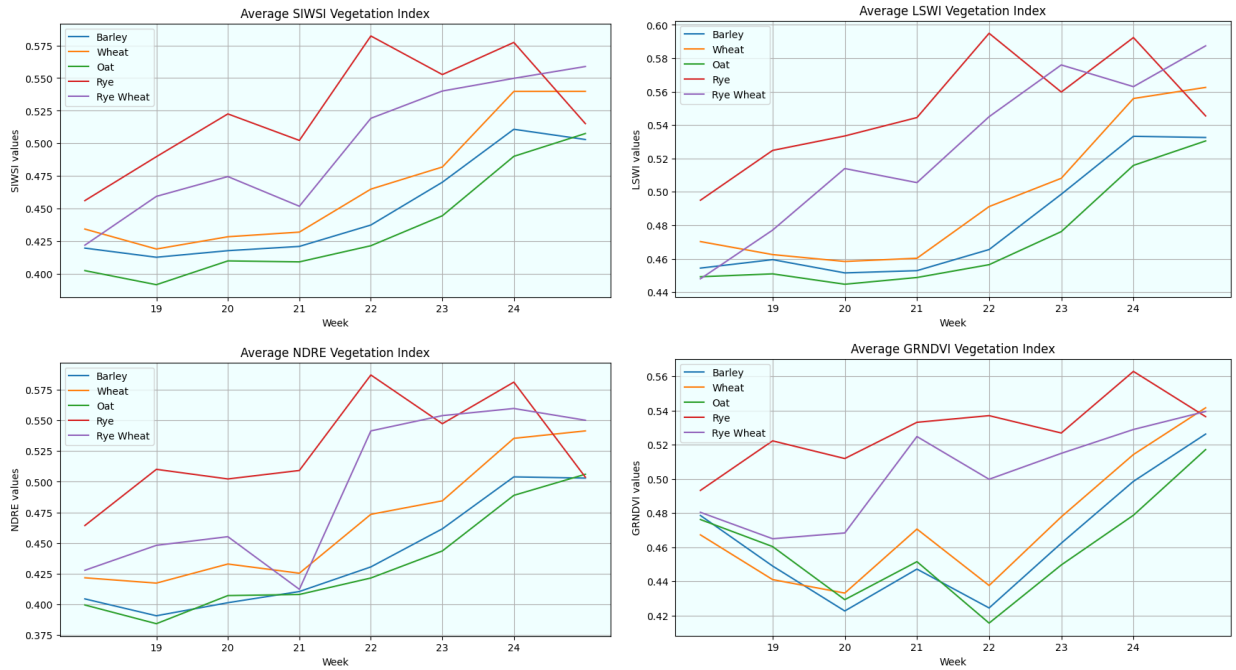

Figure 6.10: Optimal consecutive time period for grain classification, and its most distinguishable vegetation indices, SIWSI, LSWI, NDRE, and GRNDVI

contributed to zero change in the loss and accuracy, so they were excluded from this section. Figure 6.11 shows an example of how the loss and accuracy development during training of experiment 3 looked. These graphs are excluded from all other experiments, as they are all similar. None of the graphs showed anything interesting besides the fact that the classification model cannot learn the distribution of fields.

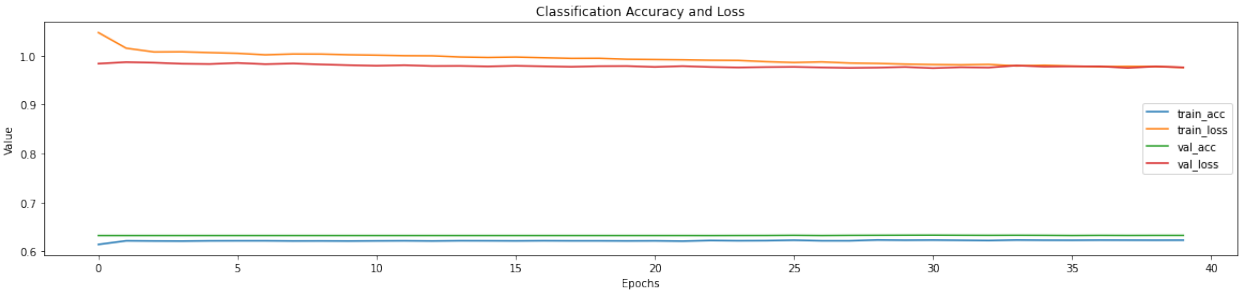

Figure 6.11: Graph showing results from vegetation indices experiment number 3

6.2.4 Multi-class Attention-based Deep Multiple Instance Learning

The MIL model performance is measured by accuracy calculated by how many bags are correctly classified as positive or negative. One model was made for each crop type because the MIL model only accepted one class as the positive one when creating the bags. Further, the model calculated an attention score during training for each sample in every bag. If a bag is classified as positive, the samples that resulted in a positively predicted bag will be given the highest attention score. Therefore each of our models will be able to classify all included crop types using these attention scores. The results for all MIL experiments are presented below using different number of bags.

Crop Type	Metric	100 bags	200 bags	300 bags	500 bags	750 bags
Barley	Accuracy	0.73	0.62	0.66	0.64	**0.74**
Barley	Loss	0.67	0.64	0.62	0.61	**0.60**
Wheat	Accuracy	**0.71**	0.70	0.73	0.67	0.68
Wheat	Loss	0.65	0.95	0.64	**0.62**	0.67
Oat	Accuracy	0.56	0.59	0.60	**0.63**	0.60
Oat	Loss	0.71	0.69	0.68	0.68	**0.67**
Rye	Accuracy	0.96	0.97	0.98	0.99	**0.99**
Rye	Loss	0.37	0.47	0.45	0.48	**0.48**
Rye Wheat	Accuracy	0.96	0.97	0.98	**0.99**	0.97
Rye Wheat	Loss	0.21	0.15	0.60	**0.09**	0.21

Table 6.3: Validation accuracies achieved from all MIL models with different number of bags

Table 6.3 shows each model's accuracy and loss results using different parameters. The text marked as bold is the best result per row. As we can see, barley, wheat, and oat have almost similar results. On the contrary, rye and rye wheat have significantly higher accuracy values and lower loss values. Additionally, better results are achieved using more bags as training data. The accuracy tends to increase, and the loss tends to decrease as we advance the number of bags. The reasoning behind these results will be discussed in the upcoming chapter.

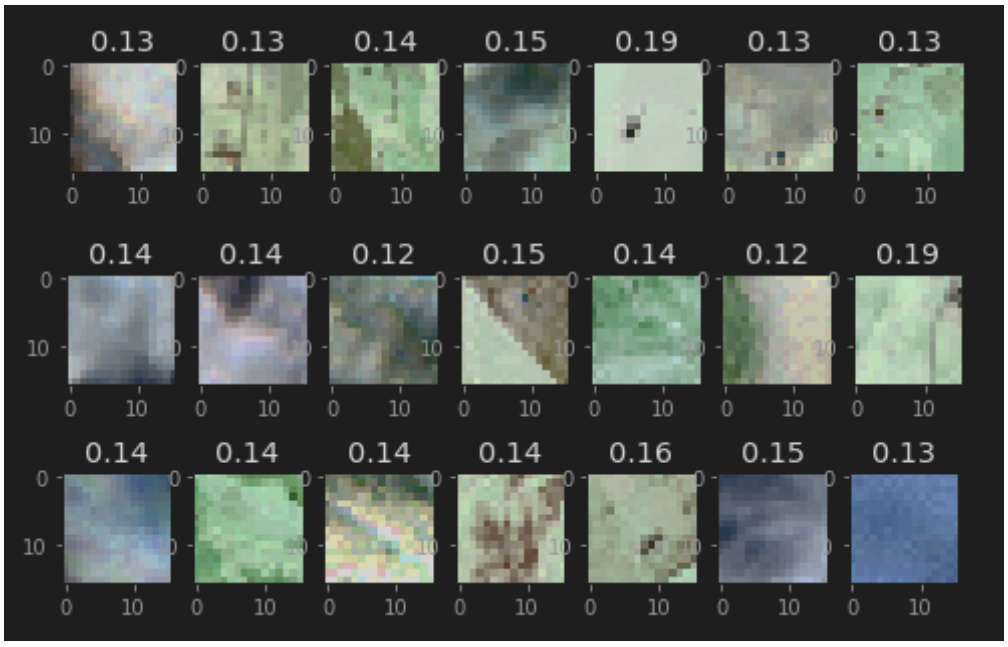

Figure 6.12: Example of predictions from MIL model showing attention scores

Figure 6.12 shows the extracted attention scores from one of our trained MIL models. This figure depicts three bags, one bag each row, using a bag size of 7. The first two rows are positively predicted bags, and the last row is negatively predicted. This particular model was trained using barley as the positive label. Additionally, we can see that some instances (field-images) have a higher score than others. The two fields that got a score of 0.19 are most likely a field containing barley. In the bottom row, the scores are slightly more equal.

6.3 Novel Approach to Early Grain Yield Predictions

As described in chapter 5, two early yield prediction and crop classification models were implemented to attempt in-season predictions only using data acquired during the growing season. This implementation uses no data acquired from delivery or farm subsidies. It has less data than the early yield prediction model from the previous students but should be more applicable for in-season predictions, which is more realistic and usable for real life situations.

6.3.1 Results on Early Crop Type Classification

Firstly, two crop classification models were implemented and trained on data in the two time periods described in chapter 5. These models were identical to the field-based classification models, using field-based satellite images as input. The only difference was the number of images fed to the model.

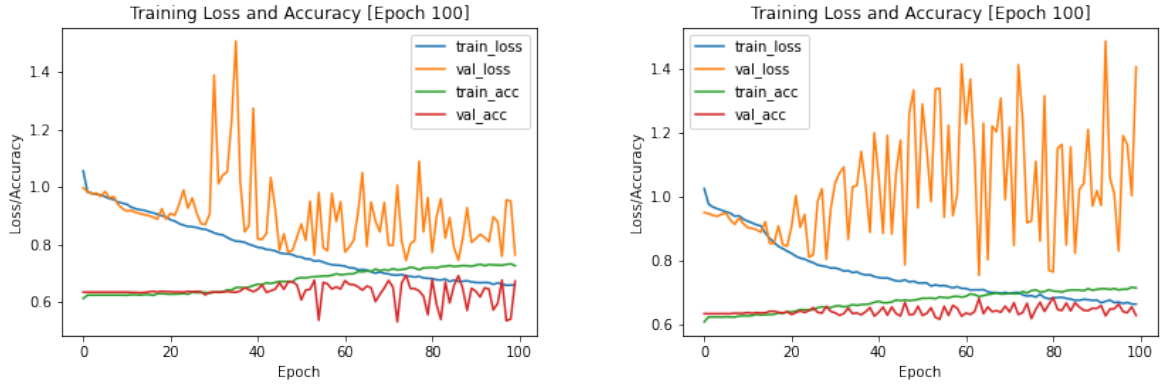

Figure 6.13: Results for the early crop type prediction model using week 10-26 (left image) and week 10-21 (right image) as data

Figure 6.13 shows loss and accuracy values for both classification models implemented. It clearly shows that the model can learn from the training data, achieving an accuracy of 70% for both models. However, as illustrated by their validation losses, the models struggle with the validation data. Despite the validation losses, the models achieve about 63% validation accuracy.

6.3.2 Results on Early Grain Yield Prediction

Further, the classification models were used to predict planted crops on all field data. These predictions were then used to create crop-specific masks per farm and feature for the early yield prediction model. Adding these additional features, including weather features and soil quality, the early yield prediction models were trained as presented in chapter 5.

Input	Description	Previous MAE	Achieved MAE	Change
Weeks 10-39	Full season	76.27 kg/daa	78.96 kg/daa	3.53%
Weeks 10-26	Late-June	82.11 kg/daa	89.77 kg/daa	-9.33%
Weeks 10-21	Mid-May	92.20 kg/daa	93.45 kg/daa	-1.35%

Table 6.4: The mean absolute error (MAE) achieved using different time periods with the hybrid early yield prediction CNN. The MAE are compared to the previously achieved values from Engen et al. [12, 13], using the same setting.

Table 6.4 shows our results from the two time intervals early in the growing season used by Engen et al. [12, 13]. As we can see, our results are comparable and not significantly worse than the previous ones. Our results were also achieved using less data and, more importantly, only data available during the growing season. These results suggest that early predictions are achievable using our data as described.

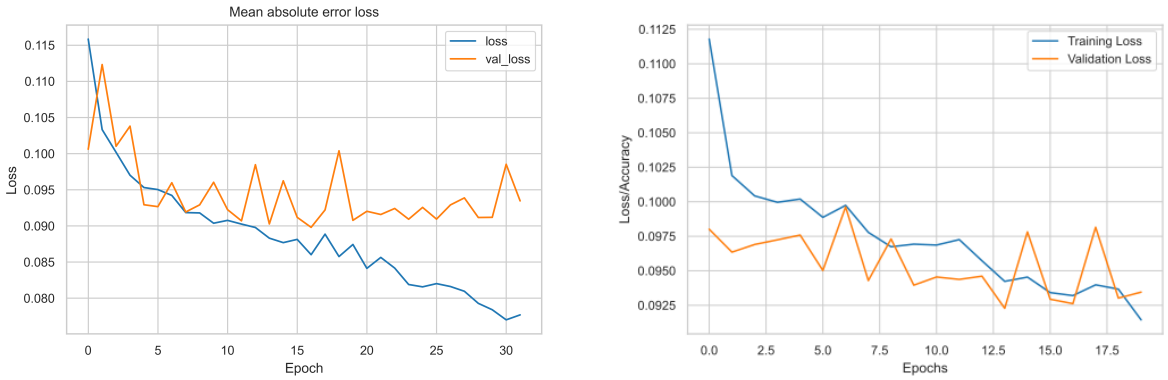

Figure 6.14: Training and validation loss for the hybrid early yield prediction model using week 10-26 (left image) and week 10-21 (right image) as data

Figure 6.11 illustrates the loss achieved by our novel approach to early yield prediction. The validation losses achieved by the week 10-26 and week 10-21 models were 0.090 and 0.093, respectively. Both models show a decrease in loss during training, which confirms that they managed to learn from our dataset. The graphs also show tendencies of overfitting, which was especially common for our data. Overfitting was an issue during most of our experiments.

6.4 Summary

The results in sections 6.1 to 6.3 show how our data was used for creating different machine learning models:

1. A yield prediction model was improved by adding more data.
2. Different crop classification models were proposed using different machine learning methods.
3. An early yield prediction was improved using the crop classification model.

Table 6.5 below shows our best results from each section split between crop classification and yield prediction.

Metric	**Yield prediction**	**Hybrid Yield Prediction**	**Early Yield Prediction**
Loss (MAE)	82.1 kg/daa	78.96 kg/daa	89.77 kg/daa

Metric	**Farm-Scale**	**Field-Scale**	**Vegetation Indices**	**MIL Classification**
Accuracy	0.60	0.70	0.64	0.74

Table 6.5: Summary of the best yield prediction (top table) results, and crop classification results (bottom table) results

Chapter 7

Discussion

The range of experiments in this research effort on applying crop classification and yield prediction has presented several interpretable results. In addition, while preparing and applying these experiments, many challenges and limitations presented themselves. The initial research challenges presented in Chapter 1 will be addressed throughout this chapter, and solutions are discussed. In addition, the results achieved from experiments will be interpreted, evaluated, and reflected. A considerable amount of slight improvement and more extensive experiments will be suggested as future work.

One of the most considerable limitations to our research effort is the lack of features related to crop growth, which has likely severely affected the machine learning performance of our experiments. Additional information such as what the farmer has planted for each field would provide valuable information for crop classification. This could be supplemented with accurate field boundaries and the activities performed on the fields, like sowing time, fertilization, irrigation, and stone picking, for more information related to growth. However, parts of this study's innovation are its unique combination of available data compared to other studies. This situation is typical for research in agriculture as there exist a vast amount of data related to crop growth. However, countries and actors usually only have the resource to record some of it. This is also affected by Norway's unique environment and climate compared to the more significant grain producers such as U.S and Ukraine, as is understood from the literature review.

One of the most considerable challenges has been the computing requirements in terms of GPU for training the networks and pure processing power when loading files into memory. Training through one epoch of the hybrid network with all features requires the computer to load, prepare and process about 250 GB of data. As this was too much to load into memory at once, the computer had to load x amount of data samples into memory consecutively. This is where we identified much of the bottleneck. Using a modern graphics card, fully training the hybrid network took an estimated five days, meaning that additionally improving the model and tuning hyperparameters was not always achievable. One improvement to this problem was to purchase a subscription to Google Colab[1] with enough resources and the ability to run scripts over a long period without interruptions. While Colab provided enough processing resources, loading data samples from disk to memory increased in time. The Colab environment had to consecutively load samples from a file stored in some remote cloud storage. This problem was never entirely solved, and it resulted in us training the networks just enough to identify the desired trends or until limited improvements were found. It is likely that with the correct system, and most preferably, enough memory, one would be able to train the networks for more prolonged, achieving improved results.

Another recurring challenge encountered during our experiments was overfitting. Almost every experiment resulted in some magnitude of overfitting, especially when using satellite images. Our

[1] https://colab.research.google.com/

theory is that the acquired satellite images contribute heavily to overfitting. One reason for this could be the number of bands used in the Sentinel 2A images, containing 12 bands, all using different resolutions. The different resolutions could lead to monotonous and minor variations in the images. Another problem with the satellite images is the amount of noise, which is discussed in section 7.4. Other reasons for overfitting could be model sizes and parameters. Even though these numbers were adjusted to minimize overfitting, it could still be a recurring problem. Especially when using the time distributed models, which leads to a bigger model containing many trainable parameters. Lastly, one reason for overfitting could be the lack of data. Despite using gigabytes of data, the models might find finding patterns and making generalizations difficult. Our results showed that adding additional data samples and features improved the model, implying that this could benefit models when nullifying overfitting.

7.1 Interpreting Baseline Improvements

Overfitting is a phenomenon that occurred throughout our research efforts. This was first identified during the recreation of the DNN by Engen et al. [12, 13]. Our research on overfitting and how to overcome this problem became very useful to us as it resulted in an improved model. Model Architecture changes and improvements were applied throughout our research efforts, and generally limited overfitting for all of our models.

Adding new meteorological features to the dense model achieved competitive performance to the state-of-the-art Engen et al. [12, 13]. However, adding these new features to the hybrid model that used satellite images showed an increase in loss and MAE. This MAE was calculated using the model's latest iteration and not the best version, as the best was not saved. This is just a mistake in our implementations, and seeing as training the model took a considerable amount of time, this was not done again. Applying the best iteration of the model would likely have resulted in an MAE similar to the previous results from Engen et al. [12, 13]. Why adding new features to the hybrid model increased the MAE is unknown. The only uniqueness of this experiment is the combination of new features with the satellite image. Therefore, the problem could be that the time distribution of the new features does not correlate with the time distribution of the images. The problem could also be related to the hybrid models architecture, whereas it might not be optimal for the extended amount of features.

While the improvements were minimal, our baseline experiments on including new features support the theory of growth factors, showing the first hypothesis to hold true to some extent. The results from our experiments indicates that grain yields are dependent on more than the features used by previous studies on yield prediction, and it is now even more likely to believe that the accuracy can be improved. Based on our research experiments and findings, features such as sowing time, field techniques, fertilization, irrigation, and others, are now more likely to increase the accuracy of yield prediction models.

Adding more data samples to the DNN model saw a decrease in MAE compared to the results from Engen et al. [12, 13]. While the decrease is lower than expected, the results are as hypothesized. Merging the experiments into a model with new features and samples also decreased the MAE of predictions. The MAE was right between the MAEs calculated using only new features and only new labels. This was expected as both the new features and the new samples improved the model independently. Adding more data samples proved to positively affect the performance of the yield prediction, which is an important finding, as new samples can be included as each growing season passes, further improving the model for each year. It would be interesting to experiment with the difference in the hybrid model when applying new data samples after five to ten growing seasons have passed. Customizing the code-base for new features and more data samples improved the system's usability and enabled the system to include additional new features and samples more conveniently.

7.2 Interpreting Results from Crop Classification

Even though the farm-scale classification models achieved a validation accuracy similar to the model that used vegetation indices, applying a farm-based crop mapping in a natural environment is unpractical. Farmers often produce different types of the crop in the same growing season. Therefore, labeling each farmer with a single type each year would be wrong. This is why we created the field-scale classification system. Through our research efforts and attained knowledge, it is in our beliefs that crop type mapping performs best when generalized across particular arable land areas, such as fields, and not farm properties. This has also proven to be true for our experiments.

By improving the classification model over time and customizing the training data to most optimally represent information on the fields, classification accuracy was improved, and overfitting was significantly reduced. Implementing the field-scale crop classification model yielded an accuracy increase of almost 10%, suggesting that field-based data is better suited for our models. Overfitting was also reduced mainly by reducing the layer sizes and removing some layers. This model reduction proves that smaller model sizes are better for reducing overfitting.

Kussul et al. [14] achieved a classification accuracy as high as 94.6%, which is a significant difference from our best model, which reached an accuracy of 70%. While they used a model architecture similar to ours, applying multiple 2D CNNs, their environment was still significantly different. Firstly, they applied pixel-based cloud and shadow removal for all satellite images. Additionally, they applied multi-source multi-temporal satellite images acquired from Landsat-8 and Sentinel-1A. These differences likely explain why they achieved such high accuracy, having access to more variety of satellite images free from clouds and shadows. If this is the case, it significantly strengthens the need for cloud and other noise removals from our satellite images.

The only known work on classifying crop types in an environment similar to ours is the application developed by Digifarm, which has achieved an accuracy of 92%. This difference is most likely due to their model using 1-meter resolution satellite images with accurate field boundaries and crop labels, while our implementation is negatively affected by the lower quality of the respective data. In addition, they applied their model to a single region in Norway, while our techniques are generalized across every single grain farmer. Introducing this country-wide generalization is likely to affect the performance, as more variety between the samples is introduced. Considering these differences, our successful research effort on crop classification can be considered a novel approach to crop mapping in Norwegian agriculture, being the first of its kind.

7.3 Interpreting Results from MIL Model Crop Classification

A novel approach to Deep Attention-based Multiple Instance Learning was developed by introducing multi-class. MIL techniques are generally designed for single-class problems, and the framework by Ilse et al. [86], which was extended for our methods, does not directly support multi-class instances. Our system was made multi-class as described in Chapter 5. Our model is not directly multi-class but requires multiple models to make classifications, which is one of the improvements that could make the model perform better. Another major limitation of the implementation is that only about 10% of our images were used because of memory restrictions and problems with TensorFlow Generators. However, the model is able to train using both labeled and unlabeled data, which is already an improvement compared to the other classification models. The fact that our MIL models perform equally, if not better than our other best crop classification models, is another considerable improvement. Giving a model more data samples has already proven to impact machine learning models positively. Additional data have already shown a decrease in the loss for all MIL models, proving our theory, but training the MIL model using even more data samples would likely have yielded better performance.

The distinction mentioned in the previous chapter, where rye's and rye wheat's results are very different, is highly correlated to the number of crop-type samples. As shown in figure 6.6, rye and rye wheat is highly underrepresented, which leads to a massive imbalance between negative and positive bags. This imbalance definitely influences these results because, in most cases, the number of positive bags for a training session was between 1 and 4. This means that the model could learn itself to predict every bag as negative and still achieve an accuracy of about 99%. However, this is not the case for barley, wheat, and oat because the bag ratio for these models was considerably more balanced.

When studying the attention scores shown in the previous chapter in figure 6.12, we did get a proper distinction between the fields in each bag. When looking at the scores, it is unclear what fields contributed to the prediction, and one can only make assumptions. A "better" MIL model would produce attention scores much more like the one shown in the work by Isle et al. [86]. Again, a reason for this could be our satellite images, which did not produce much better results in our other experiments. Another reason could be the amount of data we used, as we were limited to only 10% of our complete dataset. Image augmentation would also further extend our dataset and could have made our model better at generalizing. However, because of memory restrictions, this was inconvenient to implement. While the results are generally similar to the initial field-based approach, some novelty occurred when the state-of-the-art technique was extended to support a semi-labeled multi-class dataset.

7.4 Interpreting Results from Early Yield Predictions

Our early yield prediction experiments achieved a higher MAE than those previously achieved for the same time intervals. These results were somewhat expected, as necessary information correlated with the yield was removed while more uncertain data was included. Our goal for this experiment was never to improve early yield prediction performance in terms of accuracy and MAE. Instead, it revolved around creating a realistic and ethical system for early yield prediction, in terms of features available, without losing too much accuracy. Our main contributions to this system were the classification of crop types and the estimation of field areas, which were applied to the prediction model. The early classification models were both able to achieve an accuracy of 70% and are stable, losing minimal accuracy even when the number of features is halved. Applying the early crop classification results for the 16 first weeks of the growing season increased the MAE by 9.33%, while it only increased by 1.35% when using the first 11 weeks. This was a surprising difference, as both early classification models that predicted some yield features achieved the same performance. Therefore, we cannot conclude why the differences between these performances are as high or why the technique with more features achieved less performance.

We are generally satisfied with both early yield results. They prove that a realistic yield prediction model can be developed without losing too much performance, showing our last hypothesis to hold. To the best of our knowledge, this technique is currently the most optimal and realistic implementation of yield prediction applicable to Norwegian agriculture, being able to output predictions throughout the growing season.

A limitation of our novel approach to early yield prediction is that rye was not included. Barley, wheat and oat are the only crop types used in our novel approach. The reason behind this is that the early classification models never predicted rye, which is most likely because rye samples are significantly underrepresented in our dataset. Rye is one of the less planted grains in Norway, meaning that the dataset imbalance is hard to fix.

Several small changes can be applied in an effort to improve this accuracy while still only using available data. Currently, each sample has the historical yield for all crop types attached as features. This was done to keep our experiments as similar and consistent as Engen et al. [12, 13] by only

changing features not available at the time of the prediction. These historical features could be changed to only cover the historical yield of the target crop type, alongside the historical area used for each of these years. This inclusion should minimize disturbance from unrelated features while enabling the model to learn the correlation between total field area and crop yield. Implementing these historical features is currently achievable as the area used for each crop is only available from 2017 and onwards. While one could attempt to estimate the area used, as done in the early yield experiments, this historical implementation will become more feasible as each growing season passes. This also applies to historical subsidies, which would be an exciting set of features. Today's detailed subsidies data has only been recorded since 2017 and is therefore not yet suited as a historical feature.

Another improvement that could be made is regarding soil quality. Currently, the four soil quality features applied to the yield prediction model are estimated based on all the fields of a farmer. Instead, this should be estimated for each crop type a farmer is producing in the given year. The soil quality for a given crop would be estimated only based on the field used for that crop in that given year. Instead, it is based on all of the farmers' fields.

Our last suggestion on improving early yield predictions is related to estimating the future meteorological features. Meteorological factors have been shown throughout this research effort to have significant importance for crops to flourish and mature and therefore have a decent correlation to the yield. The accuracy is likely to drop as more and more of these features are removed the earlier one performs yield predictions. For these reasons, just as it has been explained earlier in Chapter 3, future weather predictions should be part of early yield predictions as a prerequisite task. This research effort did not include this task because of time restrictions. To process enough years of data, design a prediction model, and then evaluate the results would take a significant amount of time. This would likely have been the following experiment to perform in this research effort without the time restrictions. The Frost API has multiple years of meteorological data easily accessible, and the scripts for processing these are already developed. Therefore, it is fair to believe that predicting meteorological features is achievable to some extent. It would be fascinating to experiment with how well one can predict meteorological data in Norway, given factors like climate changes, and how well one can improve early yield predictions by applying estimations of the future weather.

7.5 Comprehending the State of Satellite Images

The satellite images acquired from KORNMO were, since the initial parts of our research efforts, known to contain some amount of clouds. It was also known that the previous research effort by Engen et al. [12, 13], who acquired the satellite images, excluded approximately the five most cloudy images. In addition, they were able to use their satellite images to predict crop yield with minimal loss. Due to these reasons, the satellite images were not validated further during our research efforts, which has later proven to be a mistake. Examples have been shown throughout this research effort, and many incidents were found during our pre-processing and experiments. Not once was a farm found with clean images throughout the whole growing season. Clouds and shadows, which are often connected, were the most occurring noise, while other types of sensor disturbance were found from time to time. Seeing as satellite images are a part of the foundation for all of our research efforts, it is highly recommended that a framework for acquiring, cleaning, and evaluating satellite images be developed before further studies using the images are done. Techniques for cloud and noise removal were presented in Chapters 2 & 3. Some of them were presented as promising techniques that were supposed to be simple to implement. However, the restricted time of our research effort prohibited us from adequately performing cloud removal.

Another problem with the current satellite images was their resolution. Sentinel-2A has bands ranging from 10 to 60 meters per pixel resolution, which severely limits how much information can be represented in each image. This was seen in the 16 x 16 field-based images, where most fields

could fit size of 10 x 10. Such resolutions are probably even less suitable in Norway, considering the geography. Norway has a lot of mountains and valleys, and the land is rarely flat over a large area. This has resulted in less arable land where fields are divided into small areas that are rarely flat. This is highly different from the U.S, where the states that produce the most grains have huge arable land areas consistently flat [11, 49].

The small resolution is even worse for the field-based images with masks applied, which only had about 5 x 5 area of value. However, this latter problem was most likely because of several different factors. While improvement in this area is significantly more limited than noise in images, Digifarm [22] was found to upscale their remote sensing data to 1 meter per pixel. It is unknown to us exactly how, but their product shows that it is possible and is probably one of the reasons why they can achieve 90% accuracy on crop mapping.

As presented in Chapter 1, acquiring new satellite images was identified as a challenge early on. After calculating the number of images required, a subscription to SentinelHub at a significant price-point was purchased. This subscription turned out to not have nearly enough processing units required to acquire all the satellite images because of how the system functions. After acquiring about 20% of the required satellite images, the downloading was halted for financial reasons. This is why no experiments use more data samples on the hybrid CNN model. Finding other methods for acquiring new satellite images was not prioritized as this problem only directly affected one of our hypotheses.

7.6 Evaluating Experiments and Interpreting Results from Vegetation Indices

Some vegetation indices were seen in Chapter 4 to contain limited to no information. In addition, our research efforts on crop classification and mapping using vegetation indices saw no improvement compared to the research efforts with raw satellite images. While optimizing the model for the classification of raw satellite images was successful, it resulted in an almost zero change in the accuracy when using vegetation indices. Even when an optimal period where the average vegetation indices could distinguish between each crop type, the classification model could not learn the distinction any better. There could be several reasons why this occurred. CNN might perform better using raw satellite images. This is because CNNs are designed to extract the interesting and relevant features from image-like data automatically and can learn the spatial and spectral context of images. Our CNN models might extract more relevant features from the raw images than derived vegetation indices.

As seen in our examples, the atmospheric noise is the more likely cause of poor performance with vegetation indices. As discussed earlier in this chapter, clouds and other noise forms frequently appear on our satellite images. This is likely to affect the calculations of vegetation indices. The idea of vegetation indices is to combine the best from different bands into a new band, which loses some value if the bands are similar because of cross-channel noise. While calculating the average vegetation indices, each image was normalized between zero and one. While doing this work, multiple images were identified to have significantly high positive and negative values, which should not be the case. Some vegetation indices likely do not account for noise in images and are therefore seriously affected by this. Foerster et al. [16] visualization of the temporal profiles, shown in Chapter 3, saw a much more consistent distinguishment between the crop types compared to our temporal visualization of NDVI in Figure 6.9. This further suggests that noise impacted the vegetation indices, as temporal noise would make the difference between each crop type smaller.

Feature selection and evaluation could have been performed within the ten selected features to evaluate how well each of the indices performed on classifying crop types. This was not done due to the time it takes to train the entire network. Based on the results from the classification

experiments, it is safe to say that raw satellite images perform better than vegetation indices in our given setting. However, considering the identified problems with the images, it is difficult to say whether raw satellite images are always more suitable than vegetation indices for grain classification.

As further work, it would be interesting to divide the Kg/daa yield for each field into one out of a few classes and then plot the average vegetation indices for each yield class. Another idea would be to evaluate the correlation between each yield and its vegetation indices values. This type of experiment could evaluate if and how well each vegetation index is on predicting the yield by checking whether the curve of the indices is different for each yield class.

7.7 Evaluating Field Boundaries

Disposed properties were not created for 2020, mainly because we could not acquire the belonging satellite images for 2020, but also because the disposed properties' shapes rarely change from one year to another. The disposed properties dataset used as field boundaries comprises a few different data sources, which are supposed to present the arable land properties of farmers. While this is true to some extent, the boundaries do not follow the actual fields accurately. If two neighboring fields are side by side, they will appear as one field in the dataset, even though farmers might dispose these fields for two completely different purposes. This is likely why our field-based masked images look weird and not like a consistent field. It is also likely one of the reasons why the field-based classification models performed better without masks. Based on the identified problems of the property and field shapes, we propose a system for field boundaries detection as future work. One could manually draw visual field boundaries on the satellite images to serve as training data and then connect these shapes to farmers based on coordinates and the existing disposed properties. A machine learning technique could then be applied to learn the distribution of field shapes and then identify field boundaries on new satellite images. More accurate field boundaries from a system like this would likely improve every experiment in this research effort where satellite images are applied.

Chapter 8

Conclusions

In conclusion, our research effort explores the use of satellite, meteorological and geographical data for improving crop yield prediction and proposing a crop classification method. New farm-based and field-based datasets have been created by combining multiple data sources, which enables the use of deep learning for crop classification and better yield prediction. To the best of our knowledge, our early yield prediction is currently the most optimal and realistic implementation of yield prediction applicable to Norwegian agriculture, being able to output predictions throughout the growing season. Multiple models are proposed and used to test different hypotheses and improve existing results within the field. The best crop classification model is a deep convolutional and recurrent model that uses multi-spectral satellite images for crop type predictions. Additionally, a novel approach for classifying semi-labeled data using multi-class attention-based deep multiple instance learning is proposed. Seeing as there exists zero previous research efforts on country-wide crop classification in Norway, Our successful research effort can be considered a novel approach.

The theory and background of the five hypotheses presented in this research effort have been researched in order to implement suitable experiments, which enabled us to evaluate and address the hypothesis as the following:

Hypothesis 1: Features related to sunlight, growth temperature, ground state, and soil quality can be applied to prediction models for more accurate grain yield predictions.

Our work on researching, exploring, and applying new data features related to plant growth enabled the yield predictor to learn the distribution between growth-related information adequately. Applying this data to the different re-implemented grain yield prediction models achieved competitive results and proved that additional growth factors could improve the system.

Hypothesis 2: Extending agricultural data for predicting grain yields with an additional year of data samples shows an increase in accuracy and the need to collect data over time.

Expanding on the code-base for feature pre-processing enabled the system to acquire data from additional growing seasons more conveniently. Applying data samples from an additional season achieved competitive results compared to the state-of-the-art by Engen et al. [12, 13], indicating that a yield prediction system like this has the potential to improve as each year passes.

Hypothesis 3: Satellite images and vegetation indices can be applied to convolutional neural networks for accurate grain classification on fields.

The results obtained from different crop classification models using multi-spectral satellite images were satisfactory when cropping the data to a more appropriate format for fields. While analyzing the temporal profiles of vegetation indices provided valuable information on the development of each crop type, the vegetation indices themselves were not suitable for crop classification in their current state.

Hypothesis 4: Multi-class Attention-based Deep Multiple Instance Learning is able to utilize the whole field dataset and thereby increase crop classification accuracy.

Our implementation of Attention-based MIL is a more suitable and accurate crop classification model for the semi-labeled field dataset by dividing fields between training bags, whether labeled or not. Given the inconclusive attention weights from the model, it is challenging to evaluate its accuracy. Therefore, we can unfortunately only conclude that our MIL model achieves comparable results.

Hypothesis 5: In-season early yield predictions are as successful when using predicted crop types limited to the data available at that time.

The best performing classification model successfully classified crop types during the growing season. These predicted crop types were successfully applied to create field masks and growing areas. Removing features unobtainable during a growing season and applying the field masks and areas to the early yield prediction model resulted in satisfactory results compared to previous work. The proposed system enables realistic in-season early predictions that could benefit actors in real-life scenarios.

8.1 Research Contributions

Our achieved contributions, justified in this research effort, are listed below.

- Improved the current system on grain yield predictions and monitoring for Norwegian farms and showed that this system now has the potential to improve as more seasons pass.
- Developed techniques for grain type classification and mapping on Norwegian fields, supporting semi-labeled fields.
- Developed an accurate, realistic, and employable system for predicting the yield of grains on Norwegian farms during the growing season.
- Identified the available data source related to growth factors in Norway, while comparing to the theory of growth factors and precision agriculture. The most significant limitation identified was the lack of data acquisition related to farmer activities.

8.2 Future Work

A range of techniques as future work were identified. While these were not implemented because of time restrictions, they are presented as highly recommended experiments in an effort to further strengthens Norway's agriculture community on crop management, monitoring and utilization.

8.2.1 Removing noise and disturbance from remote sensing

Before our acquired satellite images are further used, it is highly suggested that a system for automatically identifying and removing noise from satellite images is developed. As shown through the findings of this research effort, the quality and content of the satellite images have been identified to be insufficient. Improving the quality of images by removing noise and maybe even up-scaling resolutions have a solid potential to improve precision agriculture in Norwegian agriculture.

8.2.2 Adding additional features for yield prediction

A range of different improvements for early yield predictions has been proposed in this chapter. Data sources such as sowing area used and subsidies received should be applied as features representing historical data. The weather features removed when predicting the yield during the season should be attempted to be predicted. In addition, there is still a wide range of growth factors from Chapter 2, such as farming activities, sowing day, fertilization, and irrigation, that are likely to correlate with yields. In order to further improve yield predictions in Norwegian agriculture, one could research if and how these features could be acquired. One could also attempt to research the possibility of predicting or estimating these valuable growth factors.

8.2.3 Acquiring accurate field boundaries

As recently mentioned, we propose the work on accurate field boundaries detection. Using the already acquired data, one could manually create a training dataset for a machine learning model to detect boundaries in satellite images. While creating such a training dataset would be time-consuming, the task itself is simple. It is fair to believe that given an accurate training dataset, a machine learning model would be able to identify new boundaries given the visual differences. This technique could be connected with functions for acquiring new satellite images and removing noise from satellite images, resulting in a framework for acquiring and pre-processing images.

8.2.4 Working with government agencies

One of the most significant limitations in this research effort has been the limited amount of data related to available growth factors. To the best of our knowledge, we have explored many of the available features related to crop growth. Therefore, to further improve crop production optimization and monitoring, researchers are encouraged to work more closely with actors related to Norwegian agriculture. By working with actors such as Felleskjøpet, NIBIO, and the department of agriculture, one might initialize specific projects, which results in the collection of more agriculture data. A project like this could, for example, be to develop a system that enables farmers to continuously track how they dispose of each field and what activities they perform on each field. Not only is a research effort of this type likely to improve on the experiments in our research effort. It would also present new opportunities for other researchers to improve their experiments or perform new experiments. We suggest this future work for researchers who want to expand on the state-of-the-art to improve Norwegian agriculture's position on crop production optimization and monitoring.

Appendix A

All Average Vegetation Indices

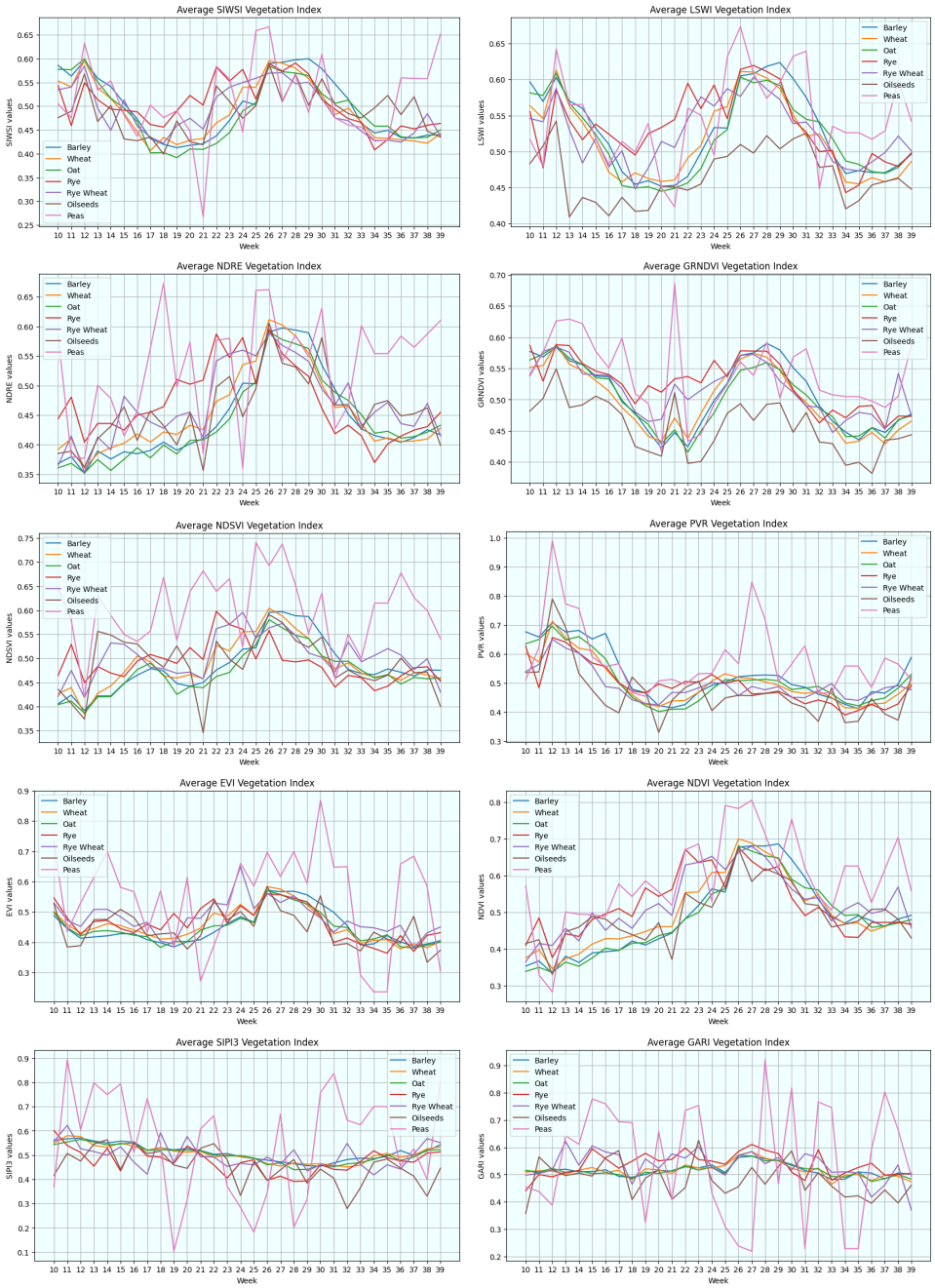

Figure A.1: Average vegetation indices for all seven crop types, throughout the growing season

Appendix B

Github Repository

The code for all our experiments and data pre-processing is uploaded to our git repository https://github.com/Ilumenar/kornmo-master-thesis. This repository is based on the work by done by Engen et al. [12, 13], therefore their repository is included as a submodule in ours. All data used in our project can be obtained from our supervisors.

Bibliography

[1] United Nations. *The 2030 Agenda and the Sustainable Development Goals: An opportunity for Latin America and the Caribbean*. (LC/G.2681-P/Rev.3), Santiago, 2018.

[2] United Nations Environment. "Towards a Green Economy: Pathways to Sustainable Development and Poverty Eradication." In: (2011), p. 52. URL: https://sustainabledevelopment.un.org/index.php?page=view&type=400&nr=126&menu=35.

[3] Jan Frouz and Jaroslava Frouzová. *Applied Ecology*. Springer Cham, 2022. ISBN: 978-3-030-83224-7. DOI: https://doi.org/10.1007/978-3-030-83225-4.

[4] United Nations Environment Programme. *Making Peace with Nature: A scientific blueprint to tackle the climate, biodiversity and pollution emergencies*. 2021. ISBN: 978-92-807-3837-7. URL: https://www.unep.org/resources/making-peace-nature.

[5] Jiaxuan You et al. "Deep gaussian process for crop yield prediction based on remote sensing data." In: *Proceedings of the Thirty-First AAAI conference on artificial intelligence*. AAAI'17. San Francisco, California, USA: AAAI Press, Feb. 2017, pp. 4559–4565.

[6] Marie Weiss, Frédéric Jacob, and Grgory Duveiller. "Remote sensing for agricultural applications: A meta-review." In: *Remote Sensing of Environment* 236 (2020). ISSN: 0034-425. DOI: https://doi.org/10.1016/j.rse.2019.111402.

[7] Gunjan Sehgal et al. "Crop Planning using Stochastic Visual Optimization." In: *2017 IEEE Visualization in Data Science (VDS)*. Phoenix, AZ, USA: IEEE, 2017, pp. 47–51. DOI: https://doi.org/10.1109/VDS.2017.8573443.

[8] Spire. *Bærekraftig Jordbruk*. URL: https://www.spireorg.no/baeligrekraftig-jordbruk.html. (accessed: 27.04.2022).

[9] Andreas Kamilaris and Francesc X. Prenafeta-Boldú. "Deep learning in agriculture: A survey." In: *Computers and Electronics in Agriculture* 147 (2018), pp. 70–90. ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2018.02.016.

[10] Sagarika Sharma, Sujit Rai, and Narayanan C Krishnan. "Wheat Crop Yield Prediction Using Deep LSTM Model." In: *CoRR* abs/2011.1498 (2020), p. 8. DOI: https://doi.org/10.48550/arXiv.2011.01498.

[11] Helena Russello and Wenling Shang. "Convolutional Neural Networks for Crop Yield Prediction using Satellite Images." In: *IBM Center for Advanced Studies* (2018).

[12] Martin Engen et al. "Farm-Scale Crop Yield Prediction from Multi-Temporal Data Using Deep Hybrid Neural Networks." In: *Agronomy* 11.2576 (2021). ISSN: 2073-4395. DOI: https://doi.org/10.3390/agronomy11122576.

[13] Martin Engen, Erik Sandø, and Benjamin Lucas Oscar Sjølander. *Deep Hybrid Neural Networks on Multi-temporal Satellite Data: Predicting Farm-scale Crop Yields*. 2020. DOI: https://hdl.handle.net/11250/2823807.

[14] Nataliia Kussul et al. "Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data." In: *IEEE Geoscience and Remote Sensing Letters* 14.5 (2017), pp. 778–782. ISSN: 1558-0571. DOI: https://doi.org/10.1109/LGRS.2017.2681128.

[15] Shunping Ji et al. "3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images." In: *Remote Sensing* 10.1 (2018). ISSN: 2072-4292. DOI: https://doi.org/10.3390/rs10010075.

[16] Saskia Foerster et al. "Crop type mapping using spectral–temporal profiles and phenological information." In: *Computers and Electronics in Agriculture* 89 (2012), pp. 30–40. ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2012.07.015.

[17] Richard Massey et al. "MODIS phenology-derived, multi-year distribution of conterminous U.S. crop types." In: *Remote Sensing of Environment* 198 (2017), pp. 490–503. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2017.06.033.

[18] Qiong Hu et al. "Integrating coarse-resolution images and agricultural statistics to generate sub-pixel crop type maps and reconciled area estimates." In: *Remote Sensing of Environment* 258 (2021), p. 14. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2021.112365.

[19] Statistics Norway. *Holdings, agricultural area and livestock.* https://www.ssb.no/en/jord-skog-jakt-og-fiskeri/jordbruk/statistikk/gardsbruk-jordbruksareal-og-husdyr and https://www.ssb.no/jord-skog-jakt-og-fiskeri/faktaside/jordbruk. (accessed: 23.03.2022).

[20] Konstantinos G Liakos et al. "Machine learning in agriculture: A review." In: *Sensors* 18.8 (2018), p. 2674. ISSN: 1424-8220. DOI: 10.3390/s18082674.

[21] Takeshi Horie, Masaharu Yajima, and Hiroshi Nakagawa. "Yield forecasting." In: *Agricultural systems* 40.1-3 (1992), pp. 211–236. ISSN: 0308-521X. DOI: https://doi.org/10.1016/0308-521X(92)90022-G.

[22] Nils Helset and Konstantin Varik. *DigiFarm.* URL: https://digifarm.io/product/. (accessed: 13.05.2022).

[23] Brett Whelan and James Taylor. *Precision Agriculture for Grain Production Systems.* CSIRO Publishing, 2013, p. 208. ISBN: 978-0-643-10749-6. DOI: 10.1071/9780643107489.

[24] Store Norske Leksikon. *Jordbruk i Norge.* URL: https://snl.no/jordbruk_i_Norge. (accessed: 01.06.2022).

[25] Hilde Halland, Mette Thomsen, and Sigridur Dalmansdottir. "Dyrking og bruk av korn i Nord-Norge. Kunnskap fra det Nord-Atlantiske prosjektet Northern Cereals 2015-2018." In: *NIBIO Rapport* 4.86 (2018), pp. 116–128. DOI: http://hdl.handle.net/11250/2504503.

[26] Ann Marie VanDerZanden. *Environmental factors affecting plant growth.* 2008. URL: https://extension.oregonstate.edu/gardening/techniques/environmental-%20factorsaffecting-plant-growth. (accessed: 02.04.2022).

[27] David M. Johnson. "An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States." In: *Remote Sensing of Environment* 141 (2014), pp. 116–128. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2013.10.027.

[28] Satish C Bhatla and Manju A Lal. *Plant physiology, development and metabolism.* Springer, 2018. ISBN: 978-981-13-2023-1. DOI: https://doi.org/10.1007/978-981-13-2023-1.

[29] Meetpal S Kukal and Suat Irmak. "Climate-Driven Crop Yield and Yield Variability and Climate Change Impacts on the U.S. Great Plains Agricultural Production." In: *Sci Rep* 8.3450 (2018). ISSN: 2045-2322. DOI: https://doi.org/10.1038/s41598-018-21848-2.

[30] NIBIO. *Økologisk landbruk.* URL: https://www.nibio.no/tema/mat/okologisk-landbruk. (accessed: 01.06.2022).

[31] Randi Berland Frøseth et al. "Våronn plan B Jordarbeiding for såing av korn under ulaglige forhold." In: *NIBIO Rapport* 6.94 (2020). ISSN: 2464-1162. DOI: https://hdl.handle.net/11250/2658924.

[32] Wajid Ali Shah et al. "Effect of Sowing Dates on the Yield and Yield Components of Different Wheat Varieties." In: *Journal of Agronomy* 5.1 (2006), pp. 106–110. ISSN: 1812-5417. DOI: 10.3923/ja.2006.106.110.

[33] Robert N Colwell. "Manual of remote sensing. Second edition." In: *Canadian Journal of Remote Sensing* (1985), p. 1310. URL: https://www.osti.gov/biblio/5772892.

[34] Isabel Luisa Castillejo-González et al. "Object- and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery." In: *Computers and Electronics in Agriculture* (2009). ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2009.06.004.

[35] The European Space Agency. *Second Copernicus environmental satellite safely in orbit.* URL: https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2/Second_Copernicus_environmental_satellite_safely_in_orbit. (accessed: 01.06.2022).

[36] Anna Chlingaryan, Salah Sukkarieh, and Brett Whelan. "Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture: A Review." In: *Computers and Electronics in Agriculture* 151 (2018), pp. 61–69. ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2018.05.012.

[37] Verena Henrich et al. *Indices for Sentinel-2A.* URL: https://www.indexdatabase.de/db/is.php?sensor_id=96. (accessed: 11.05.2022).

[38] Huanfeng Shen et al. "Missing Information Reconstruction of Remote Sensing Data: A Technical Review." In: *IEEE Geoscience and Remote Sensing Magazine* 3.3 (2015), pp. 61–85. ISSN: 2168-6831. DOI: https://doi.org/10.1109/MGRS.2015.2441912.

[39] Zhe Zhu and Curtis E Woodcock. "Object-based cloud and cloud shadow detection in Landsat imagery." In: *Remote sensing of environment* 118 (2012), pp. 83–94. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2011.10.028.

[40] Junchang Ju and David P Roy. "The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally." In: *Remote Sensing of Environment* 112.3 (2008), pp. 1196–1211. ISSN: 0034–4257. DOI: https://doi.org/10.1016/j.rse.2007.08.011.

[41] Sergey V Skakun and Ruslan M Basarab. "Reconstruction of Missing Data in Time-Series of Optical Satellite Images Using Self-Organizing Kohonen Maps." In: *Journal of Automation and Information Sciences* 46.12 (2014), pp. 19–26. ISSN: 1064-2315. DOI: 10.1615/JAutomatInfScien.v46.i12.30.

[42] Qiang Zhang et al. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 162 (2020), pp. 148–160. ISSN: 0924-2716. DOI: https://doi.org/10.1016/j.isprsjprs.2020.02.008.

[43] Chao-Hung Lin et al. "Cloud Removal From Multitemporal Satellite Images Using Information Cloning." In: *IEEE Transactions on Geoscience and Remote Sensing* 51.1 (2012), pp. 232–241. ISSN: 1558-0644. DOI: 10.1109/TGRS.2012.2197682.

[44] Qiang Zhang et al. "Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network." In: *IEEE Transactions on Geoscience and Remote Sensing* 56.8 (2018), pp. 4274–4288. ISSN: 0196–2892. DOI: https://doi.org/10.1109/TGRS.2018.2810208.

[45] Abdou Bannari et al. "A Review of Vegetation Indices." In: *Remote Sensing Reviews* 13.1–2 (1995), pp. 95–120. ISSN: 0275-7257. DOI: 10.1080/02757259509532298.

[46] Alfredo Huete et al. "Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices." In: *Remote Sensing of Environment* 83.1-2 (2002), pp. 195–213. ISSN: 0034-4257. DOI: 10.1016/S0034-4257(02)00096-2.

[47] Robert Lawrence Pearson and Lee Durward Miller. "Remote Mapping of Standing Crop Biomass for Estimation of the Productivity of the Shortgrass Prairie." In: *Proceedings of the 8th International Symposium on Remote Sensing of the Environment II* (1972), pp. 1355–1379.

[48] Xiangming Xiao et al. "Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images." In: *Remote sensing of Environment* 100.1 (2006), pp. 95–113. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2005.10.004.

[49] Liheng Zhong, Peng Gong, and Gregory S Biging. "Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery." In: *Remote Sensing of Environment* 140 (2014), pp. 1–13. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2013.08.023.

[50] Nobuyuki Kobayashi et al. "Crop Classification Using Spectral Indices Derived from Sentinel-2A Imagery." In: *Journal of Information and Telecommunication* 4.1 (2020), pp. 67–90. DOI: 10.1080/24751839.2019.1694765.

[51] Jiaguo Qi et al. "RANGES Improves Satellite-based Information and Land Cover Assessments in Southwest United States." In: *Eos, Transactions American Geophysical Union* 83.51 (2002), pp. 601–606. DOI: https://doi.org/10.1029/2002EO000411.

[52] John W Rouse Jr et al. "Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation." In: *NASA/GSFCT Type III Final Report* (1973). URL: https://ntrs.nasa.gov/api/citations/19750020419/downloads/19750020419.pdf.

[53] Compton J Tucker. "Red and Photographic Infrared Linear Combinations for Monitoring Vegetation." In: *Remote Sensing of Environment* 8.2 (1979), pp. 127–150. ISSN: 0034-4257. DOI: https://doi.org/10.1016/0034-4257(79)90013-0.

[54] Rasmus Fensholt and Inge Sandholt. "Derivation of a Shortwave Infrared Water Stress Index from MODIS Near- and Shortwave Infrared Data in a Semiarid Environment." In: *Remote Sensing of Environment* 87.1 (2003), pp. 111–121. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2003.07.002.

[55] Fu-Min Wang et al. "New Vegetation Index and Its Application in Estimating Leaf Area Index of Rice." In: *Rice Science* 14.3 (2007), pp. 195–203. ISSN: 1672-6308. DOI: https://doi.org/10.1016/S1672-6308(07)60027-4.

[56] Razieh Barzin et al. "Use of UAS Multispectral Imagery at Different Physiological Stages for Yield Prediction and Input Resource Optimization in Corn." In: *Remote Sensing* 12.15 (2020). ISSN: 2072-4292. DOI: 10.3390/rs12152392.

[57] Chanreaksa Chea et al. "Optimal Models Under Multiple Resource Types for Brix Content Prediction in Sugarcane Fields using Machine Learning." In: *Remote Sensing Applications: Society and Environment* 26 (2022). ISSN: 2352-9385. DOI: https://doi.org/10.1016/j.rsase.2022.100718.

[58] J Penuelas, Frédéric Baret, and I Filella. "Semi-empirical Indices to Assess Carotenoids/chlorophyll a Ratio From Leaf Spectral Reflectance." In: *Photosynthetica* 31.2 (1995), pp. 221–230.

[59] Anatoly A Gitelson, Yoram J Kaufman, and Mark N Merzlyak. "Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS." In: *Remote Sensing of Environment* 58.3 (1996), pp. 289–298. ISSN: 0034-4257. DOI: https://doi.org/10.1016/S0034-4257(96)00072-7.

[60] Frank Rosenblatt. "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6 (1958), p. 386. DOI: https://psycnet.apa.org/doi/10.1037/h0042519.

[61] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. International series of monographs on physics. MIT press, 2016. ISBN: 9780262337373.

[62] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library." In: *Advances in Neural Information Processing Systems*. Ed. by H. Wallach et al. Vol. 32. Curran Associates, Inc, 2019, p. 12. URL: https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf.

[63] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization." In: *3rd International Conference on Learning Representations, ICLR 2015*. 2015.

[64] Tensorflow. *Dense Layer*. URL: https://keras.io/api/layers/core_layers/dense/. (accessed: 02.06.2022).

[65] Tensorflow. *Dropout Layer*. URL: https://keras.io/api/layers/regularization_layers/dropout/. (accessed: 02.06.2022).

[66] Tensorflow. *Flatten Layer*. URL: https://keras.io/api/layers/reshaping_layers/flatten/. (accessed: 02.06.2022).

[67] Tensorflow. *Batch Normalization Layer*. URL: https://keras.io/api/layers/normalization_layers/batch_normalization/. (accessed: 02.06.2022).

[68] Suvrit Sra, Sebastian Nowozin, and Stephen J Wright. *Optimization for machine learning*. Neural information processing series. Mit Press, 2012. ISBN: 9780262016469.

[69] Aurélien Géron. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligen system. eng. Second Edition*. O'Reilly, 2019. ISBN: 9781492032649.

[70] Anh H. Reynolds. *Convolutional Neural Networks (CNNs)*. URL: https://anhreynolds.com/blogs/cnn.html. (accessed: 01.06.2022).

[71] Nicole Laskowski. *recurrent neural networks*. URL: https://www.techtarget.com/searchenterpriseai/definition/recurrent-neural-networks. (accessed: 02.06.2022).

[72] Gabriel Loye. *Gated Recurrent Unit (GRU) With PyTorch*. URL: https://blog.floydhub.com/gru-with-pytorch/. (accessed: 02.06.2022).

[73] Kyunghyun Cho et al. "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation." In: *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*. 2014, pp. 1724–1734. ISBN: 978-193728496-1. DOI: 10.3115/v1/d14-1179.

[74] William R Raun et al. "In-Season Prediction of Potential Grain Yield in Winter Wheat Using Canopy Reflectance." In: *Agronomy Journal* 93.1 (2001), pp. 131–138. DOI: https://doi.org/10.2134/agronj2001.931131x.

[75] Petteri Nevavuori, Nathaniel Narra, and Tarmo Lipping. "Crop Yield Prediction with Deep Convolutional Neural Networks." In: *Computers and Electronics in Agriculture* 163 (2019). ISSN: 0168-1699. DOI: https://doi.org/10.1016/j.compag.2019.104859.

[76] Saeed Khaki and Lizhi Wang. "Crop Yield Prediction Using Deep Neural Networks." In: *Frontiers in Plant Science* 10 (2019), p. 621. DOI: https://doi.org/10.3389/fpls.2019.00621.

[77] Leo Breiman et al. *Classification and regression trees*. Routledge, 1984. ISBN: 9781315139470. DOI: https://doi.org/10.1201/9781315139470.

[78] Robert Tibshirani. "Regression Shrinkage and Selection Via the Lasso." In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (1996), pp. 267–288. DOI: https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[79] Prakash Basnyat et al. "Optimal time for remote sensing to relate to crop grain yield on the Canadian prairies." In: *Canadian Journal of Plant Science* 84.1 (2004), pp. 97–103. DOI: https://doi.org/10.4141/P03-070.

[80] Qiong Hu et al. "Extending the pairwise separability index for multicrop identification using time-series modis images." In: *IEEE Transactions on Geoscience and Remote Sensing* 54.11 (2016), pp. 6349–6361. DOI: 10.1109/TGRS.2016.2581210.

[81] Bingwen Qiu et al. "Automatic and adaptive paddy rice mapping using Landsat images: Case study in Songnen Plain in Northeast China." In: *Science of the Total Environment* 598 (2017), pp. 581–592. ISSN: 0048-9697. DOI: https://doi.org/10.1016/j.scitotenv.2017.03.221.

[82] José M Peña-Barragán et al. "Object-based crop identification using multiple vegetation indices, textural features and crop phenology." In: *Remote Sensing of Environment* 115.6 (2011), pp. 1301–1316. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2011.01.009.

[83] Thomas G Dietterich, Richard H Lathrop, and Tomás Lozano-Pérez. "Solving the multiple instance problem with axis-parallel rectangles." In: *Artificial intelligence* 89.1-2 (1997), pp. 31–71. ISSN: 0004-3702. DOI: https://doi.org/10.1016/S0004-3702(96)00034-3.

[84] Oded Maron and Tomás Lozano-Pérez. "A Framework for Multiple-Instance Learning." In: *Advances in Neural Information Processing Systems*. Vol. 10. MIT Press, 1997, pp. 570–576.

[85] Gwenolé Quellec et al. "Multiple-Instance Learning for Medical Image and Video Analysis." In: *IEEE Reviews in Biomedical Engineering* 10 (2017), pp. 213–234. ISSN: 1941-1189. DOI: https://doi.org/10.1109/RBME.2017.2651164.

[86] Maximilian Ilse, Jakub Tomczak, and Max Welling. "Attention-based Deep Multiple Instance Learning." In: *Proceedings of the 35th International Conference on Machine Learning*. Vol. 80. Proceedings of Machine Learning Research. PMLR, July 2018, pp. 2127–2136.

[87] Veronika Cheplygina, David M J Tax, and Marco Loog. "Multiple Instance Learning with Bag Dissimilarities." In: *Pattern Recognition* 48.1 (2015), pp. 264–275. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2014.07.022.

[88] Yixin Chen, Jinbo Bi, and James Ze Wang. "MILES: Multiple-Instance Learning via Embedded Instance Selection." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.12 (2006), pp. 1931–1947. ISSN: 1939-3539. DOI: https://doi.org/10.1109/TPAMI.2006.248.

[89] Jan Ramon and Luc De Raedt. "Multi Instance Neural Networks." In: *Proceedings of the ICML-2000 workshop on attribute-value and relational learning*. 2000, pp. 53–60.

[90] Vikas C Raykar et al. "Bayesian Multiple Instance Learning: Automatic Feature Selection and Inductive Transfer." In: *Proceedings of the 25th International Conference on Machine learning*. ICML 08. Helsinki, Finland: Association for Computing Machinery, 2008, pp. 808–815. ISBN: 978-160558205-4. DOI: https://doi.org/10.1145/1390156.1390258.

[91] Fei Wen et al. "Two-Pass Robust Component Analysis for Cloud Removal in Satellite Image Sequence." In: *IEEE Geoscience and Remote Sensing Letters* 15.7 (2018), pp. 1090–1094. ISSN: 1558-0571. DOI: https://doi.org/10.1109/LGRS.2018.2829028.

[92] PH Crown. "Crop Identification in a Parkland Environment Using Aerial Photography." In: *Canadian Journal of Remote Sensing* 5.2 (1979), pp. 128–135. DOI: https://doi.org/10.1080/07038992.1979.10854991.

[93] Huiqin Ma et al. "Integrating Early Growth Information to Monitor Winter Wheat Powdery Mildew Using Multi-Temporal Landsat-8 Imagery." In: *Sensors* 18.10 (2018). ISSN: 14248220. DOI: https://doi.org/10.3390/s18103290.

[94] Tri Muji Susantoro et al. "Selection of Vegetation Indices for Mapping the Sugarcane Condition Around the Oil and Gas Field of North West Java Basin, Indonesia." In: *IOP Conference Series: Earth and Environmental Science*. Vol. 149. IOP Publishing, May 2018, pp. 4559–4565. DOI: 10.1088/1755-1315/149/1/012001.

[95] Graciela Metternicht. "Vegetation Indices Derived from High-Resolution Airborne Videography for Precision Crop Management." In: *International Journal of Remote Sensing* 24.14 (2003), pp. 2855–2877. ISSN: 0143-1161. DOI: 10.1080/01431160210163074.

[96] Sentinel. *Sentinel-2*. URL: https://sentinel.esa.int/web/sentinel/missions/sentinel-2. (accessed: 04.05.2022).

[97] *The Brønnøysund Register Center*. URL: https://www.brreg.no/en/about-us-2/. (accessed: 26.05.2022).

[98] *Mapbox & Mapbox Documentation*. https://docs.mapbox.com/ and https://www.mapbox.com/. (accessed: 26.05.2022).

[99] Norwegian Institute of Bioeconomy Research. *Jordsmonnkart*. URL: https://www.nibio.no/tema/jord/jordkartlegging/jordsmonnkart. (accessed: 10.02.2022).

[100] Eumetsat Vocabulary Manager. *State of ground (with or without snow) class descriptions*. URL: https://vocabulary-manager.eumetsat.int/vocabularies/BUFR/WMO/2/TABLE_CODE_FLAG/020062. (accessed: 01.04.2022).

[101] Patrice Ferlet. *Training a neural network with an image sequence*. URL: https://medium.com/smileinnovation/training-neural-network-with-image-sequence-an-example-with-video-as-input-c3407f7a0b0f. (accessed: 01.06.2022).