UNIVERSITY OF AGDER

Neural Network for Grain Yield Predictions in Norwegian Agriculture

Benjamin Sjølander
Erik Sandø
Martin Engen

SUPERVISORS
Morten Goodwin
Simon Arenberg

IKT442 ICT Seminar 3 Project
University of Agder, 2020
Faculty of Engineering and Science
Department of ICT

UiA
University of Agder
Master's thesis

Faculty of Engineering and Science
Department of ICT

Abstract

Grain production has many contributing factors tied to how well a crop will perform over a season. With an increased amount of data available, there is potential for machine learning techniques to be applied. As part of a greater collaborative effort between Felleskjøpet and UiA, this project explores how neural networks can be used together with Norwegian agricultural data to understand better and predict grain crops' yield.

This project had four categories of data: Annual grants of the individual farmers, how much grain the farmers delivered, the geographical location of the farms, and weather. After preprocessing and filtering the data, a simple DNN model was trained to predict the relative yield (kg per hectare) per farmer. The trained DNN model could predict yield with an average error of 930 kg per hectare, or 24% of the average yield per hectare in the dataset (3828 kg). Further experiments were performed using the same model. They include comparing yields given different types of grain and substituting the weather of one year with another to evaluate its impact on overall yield.

Table of Contents

Glossary

ANN Artificial Neural Network. 2, 10

API Application Programming Interface. 14

CNN Convolutional Neural Network. 10, 34

DNN Deep Neural Network. 10, 34

GDD Growing Degree Days, used to estimate plant growth. 10

LSTM Long-Short-Term Memory, a type of recurrent neural network. 10, 11

MapBox A provider of geolocation and map services and APIs. 14

RMS Root Mean Square. 10

RMSE Root Mean Square Error. 11

RNN Recurrent Neural Network. 11

SSB Statistisk sentralbyrå/Statistics Norway. 17

List of Figures

List of Tables

Part I

Research Overview

Chapter 1

Introduction

The last decade has seen a steady decline in both the number of active grain farmers and the total landmass area used for grain farming [28]. However, at the same time, grain farmers' efficiency has increased, leading to a steady total grain production nationally. In 2019, the Norwegian farmers cooperative, Felleskjøpet (FK), received a large investment into a new project that aims to increase quality and efficiency in Norwegian grain production by combining traditional farming with new technologies for data harvesting and machine learning [1].

Farmers strive for the highest possible yields when harvesting their crops. However, the yields vary with many factors such as weather, current and previous crop types, and fertilizer, according to a survey on Norwegian oat farmers [18]. Exactly half of the asked farmers listed weather and climate effects as the primary challenge for increasing yield. A better understanding of these factors' interaction and effects may improve yields, as farmers can affect some of them, like crop type, fertilizer use, watering, and so on. While attempts at predicting crop yield with Artificial Neural Networks (ANN) have been made before [32], data for Norwegian agriculture has recently improved in terms of availability and granularity, e.g., grant applications[2]. With this in mind, we have gathered and preprocessed relevant public data and trained a simple ANN model to predict the yield of Norwegian grain farmers.

1.1 Project definition

1.1.1 Project Goals

Goal 1: Use currently available agricultural data such as how much grain a farmer delivered and how much area was utilized together with weather data with neural networks to predict grain yields for farms in Norway.

Goal 2: Use a trained model to evaluate the performance of farmers.

1.2 Project outline

This project focused on the task of predicting grain yields for farmers across Norway. The primary purpose was to see which data was available and evaluate how it could be utilized in a machine learning method such as neural networks. Much of the work went into collecting data such as weather and geo-locations of the farms and understanding the grant applications data set.

Chapter 2

Background

2.1 Felleskjøpet

Farm co-ops (member-owned organizations) have become standard practice in Norway, and these farm co-ops seek to support their members and facilitate platforms for areas such as sales, purchasing, and breeding. Felleskjøpet is one such farm co-op and consists of two regional co-ops: Felleskjøpet Agri (FKA) and Felleskjøpet Rogaland Agder (FKRA), which together covers Norway [12]. Felleskjøpet is Norway's single largest producer of feed concentrates, with a market share of 48 percent; felleskjøpet makes feed concentrates for ruminants, swine, poultry, horses, fur-bearing animals, dogs, and cats [12]. The main ingredient in feed concentrates are grain, and felleskjøpet has established itself as a significant buyer in the grain market. As a result of felleskjøpets position in the grain market, it has, on behalf of the government and the farmers, the responsibility to regulate the Norwegian grain market as a whole [12]. In addition to grains and food concentrates, felleskjøpet also has a significant role in other farm-related areas, such as farm implementations and fertilizer sales.

2.2 Norwegian agriculture

Norwegian agriculture has traditionally been family farming. With the support of society and politicians, the goal is to reach national self-sufficiency based on the available natural resources [30]. To keep agriculture active and profitable throughout the country, subsidy rates are designed to compensate for disadvantages related to topography and farm sizes. For example, land payments are differentiated by geography and type of agricultural production [8].

Although Norwegian agriculture has consisted of smaller family farms spread out, there has been a decline in the number of agricultural holdings by 50 percent in the last 30 years. Simultaneously, there has been an increase in farms' average size, from 14.7ha in 1999 to 24.7ha in 2018 [20]. The most produced categories of food throughout the country include milk and milk products, meat, poultry, eggs, potatoes, and grains [30] [27].

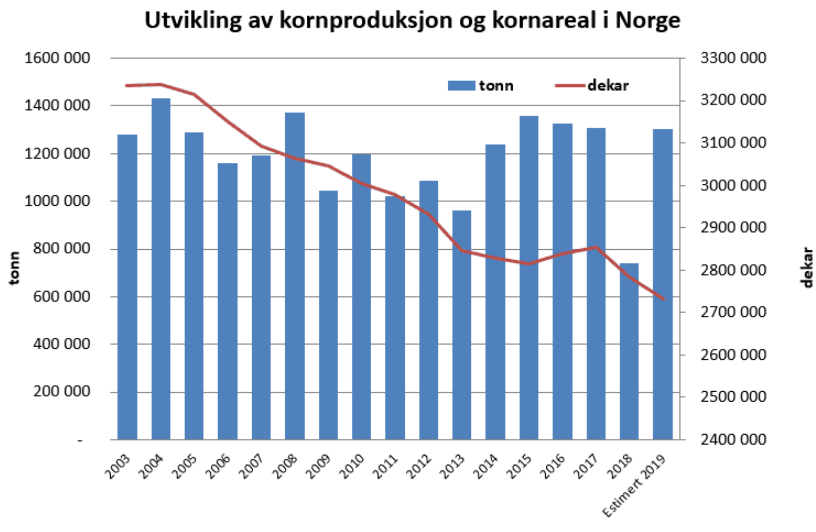


Figure 2.1: Development of grain production and acres used in Norway from 2003 to 2019.

Source: Norwegian grain production [22]

2.2.1 Norwegian grain production

The Norwegian topography consists mostly of mountain masses, and as a result, only 3 percent of the total landmass is farmed land (excluding Svalbard and Jan Mayen). Due to differences in climatic conditions, a smaller part of this farmed land can grow cereal for human consumption [20]. Using climactic divisions by Skjelvåg [26], Nibio has constructed a map to showcase where the different climatic zones are located. See appendix B for details.

As shown in appendix B, the eastern and southeastern part of Norway is best suited to produce food-grade grains. There are mainly four types of grain produced in Norway: Wheat, Barley, Oat and Rye (and rye-wheat). See the distribution of each type produces in 2019 in figure 2.2.

Barley is the most grown grain. It needs a shorter growth season before it is ready for harvest; it is well suited to be grown in areas further north and in higher altitudes areas. Most of the grown barley is used for animal fodder [3].

Wheat is a close second in terms of sum produced in Norway and requires a longer growing season than barley. Wheat can generally be split up into two categories: spring wheat and winter wheat. Winter wheat is sown in the fall, where it grows a little before the winter sets in and the plant goes dormant. Winter wheat will again resume growing in the spring and is harvested in the summer. The result is the same type of grain, but winter wheat may give higher yields as it will resume growing earlier in the spring compared to spring wheat [14].

Oats thrive in cold and moist climates, which makes them well suited for cultivation in Norway. The vast majority (more than 90 percent) of oats grown are used for animal feed, and two percent are used for human consumption [4].

Rye is the least grown grain in Norway, covering just two percent of the total areal used for grains [5]. Rye generally thrives in higher altitudes but is for the most part only grown in the east and southeastern parts of the country [5] [10].

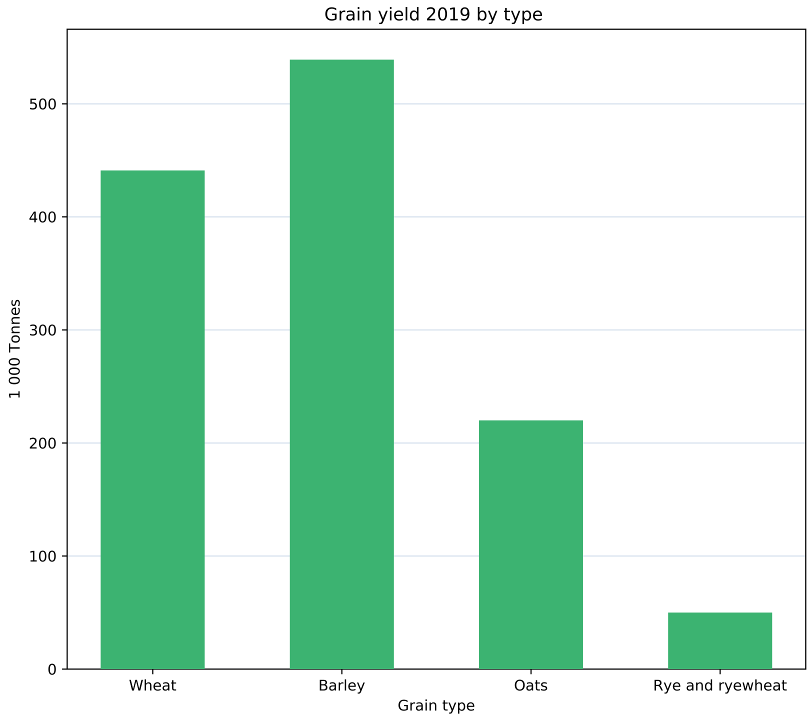


Figure 2.2: Grain production in Norway 2019 [28]

2.3 Growth factors

Plant growth is very complex, and many elements contribute to plants' growth and wellness. Species have different requirements and levels of hardiness (The ability to withstand cold temperatures). Grain plants belong to the Poaceae family, which means that it is a type of grass. Despite grass being one of the most widespread and abundant plants globally, [17] environmental factors play a crucial role in these plants' development and growth.

According to Oregon State University, four main environmental factors affect plant growth: light, temperature, water, and nutrition [31].

2.3.1 Environmental factors

Light

Light is essential for plant growth, as it is an ingredient in photosynthesis [1]. Concerning grain growth in Norway, the duration of light is especially relevant. According to Åssveen and Abrahamsen, the duration of light in a day (day length) is more influential than temperature as growth factors [33]. Day length can be calculated using the latitude of the observer.

Temperature

Temperature affects plants' growth in many ways. Perhaps most importantly, it influences the germination process to make the seeds sprout and start the initial growth. The temperature will also affect when the winter wheat breaks dormancy to resumes the growth early in spring. Most cultivated plants in Norway have a base temperature [2] of 5°C [11]. As seen in table 2.1, the types of grain have different requirements in how many sum degrees are needed before it is ripe or ready for harvest.

Growth	**Sum day degrees**
Early barley, maturation	1200
Wheat and oat, maturation	1600
Grass for feed (before first harvest)	750-800

Table 2.1: Sum degrees (in Celsius) requirements before ripe or harvest ready [11].

Water

In addition to light, water is a primary component of photosynthesis, and therefore crucial for the ultimate growth. Water can come in the form of direct precipitation or from humidity [31]. Inadequate water supply results

[1] Photosynthesis is the metabolic process of plants

[2] Base temperature is the minimum required temperature for a plant to grow. No growth happens if the temperature is below the base temperature.

in low growth and reduced yield of crops. A study conducted on Nepal crop growth concerning temperature and precipitation saw that districts that received less than half of average rainfall in the cropping season for wheat also saw crop yields reduced by more than half [7].

Nutrition

Plants need access to basic chemical elements to be able to grow. In total, 17 elements are needed for normal growth, of which three (carbon, hydrogen, and oxygen) can be found in air and water while the rest are found in soil. The roots absorb 98% of the plants' nutrients from soil-water, and if a plant is under stress by extreme temperatures, drought, or low light, it can lower the plant's ability to absorb nutrients [31]. Fertilizing is the process of adding materials that contain plant nutrients to the soil so that all nutrients are available.

2.3.2 Crop rotation

Healthy soil depends on more than just having all nutrients available, and utilizing crop-rotation can significantly impact crop yields. Crop rotation is the process of rotating what sort of crop is grown on a field from season to season. By rotating crops, it creates an opportunity to create diversity on a field [13]. Bullock shows an example of a 2-year crop rotation between maize and soybean, which resulted in 5 to 20% more maize yields. He argues that no fertilizer or pesticide can compensate entirely for that difference [9]. Crop rotations will also help break disease cycles, and including plants in the legume family in a rotation will help pull nitrogen from the air and lessen the need to fertilize nitrogen in the field [13].

Chapter 3

State-of-the-art

The use of artificial neural networks for crop yield prediction is not new; Lieu et al. published their paper in 1999, showing that ANNs can capture the nonlinear function for crop yield [21]. More recently, there have been numerous studies using Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Long-Short-Term Memory (LSTM), and other hybrid networks to predict crop yield [32]. The three most widely used NN techniques in recent publications are CNNs, LSTMs, and DNNs, in that order [32].

An early example of Artificial Neural Networks (ANN) used in crop yield prediction was performed by Lieu et al. [21] in 1999. They used a simple artificial neural network with three fully connected layers to predict corn yield production using soil data, weather data, and management data. The Morrow Plots, an experimental agricultural field located at the University of Illinois, provided the dataset used to train the ANN. The field is divided into multiple smaller plots with different plants, which are again divided into subplots, receiving different fertilizer treatments. The complete dataset for the model consisted of only 360 examples collected from 30 years of experiments. The weather data includes rainfall for each month during the season and the previous year and the calculated GDD (Growing-degree-day) for the season. Genetic data and fertilizer, planting density, and rotation factor constituted the management data. In total, the network was fed with only 15 input variables and had 20 units in the hidden layer. The ANN managed to predict the yield of 60 validation samples with an RMS error of around 20 percent[1].

Similar publications include the research by Črtomir et al., who used ANN along with image analysis for Apple yield prediction [34] by looking at the predicted number of fruits at four or five different stages of growth. By looking at maize production in east-central Indiana, USA, using data from 1901 to 1996, O'Neal et al. [23] managed to use an ANN to predict the maize yield with an RMS error of 10.5%[1], which outperformed linear and quadratic regression models used on the same dataset. They also experimented with different coding schemes for the input data, which consisted of precipitation and air temperature. They found that the neural network performed better using a max-min coding scheme (linear scaling of input values between a maximum and minimum value) compared to unary, binary, or logarithmic coding schemes.

In January 2020, Saeed Khaki, Lizhi Wang, and Sotirios Archontoulis published their research using convolutional and recurrent neural networks together to predict crop yield of corn (maize) and soybean farms [19]. Using data from the Corn Belt, across 13 different states in the United States, for the years 2016, 2017, and 2018, their model managed to predict the crop yield with an RMSE of 9% on corn yield and 8% on soybean yield[1], based on environmental and management data. Their model uses one-dimensional convolution on weekly averaged weather data and one-dimensional convolution on soil data captured at nine different depths underground. The output of the two CNNs is then fed through densely connected layers and given as input to an LSTM RNN.

[1] It is difficult to determine exactly how the different authors have calculated RMS in the form of a percentage or if they are using the same formula. The values presented here are the values the authors have used themselves.

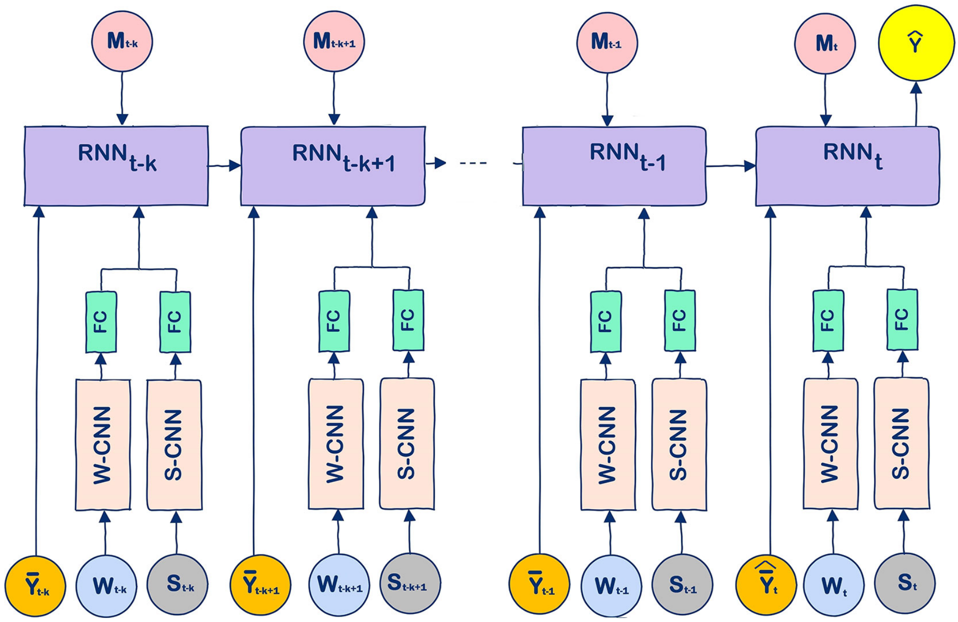


Figure 3.1: The proposed CNN-RNN model by Khaki et al. [19].

W-CNN and S-CNN are normal CNNs, with one-dimensional convolution, for weather data and soil data, respectively. W, S, and M inputs are weather data, soil data, and management data, respectively.

Chapter 4

Method

4.1 Data

Several data sets from different sources were available through various APIs and online services. These were linked by using the farmers' organization number, location, and time.

4.1.1 Grant Applications

The farmers can (and almost universally do) apply for grants from the Norwegian Agriculture Agency. We have used this application data to retrieve which grain types a farmer has grown and on how much land.

4.1.2 Grain Deliveries

The grain deliveries from various farmers describe the number of kilos of grain per delivery. There is some uncertainty to the numbers' accuracy, as one farmer can deliver on behalf of another and settle among themselves. The deliveries are split per type of grain. As not all the grain from a farmer is delivered at once, we aggregate a farmer's deliveries, resulting in the sum per grain type per year.

4.1.3 Geographical location

The address(es) for each organization were gathered from the Brønnøysund Register Centre's Unit Register API. The addresses' global coordinates were obtained by querying a Geocoding API from MapBox. These coordinates were subsequently used to query a different MapBox API to retrieve the elevation of each address.

4.1.4 Weather

While there is no weather information measured per farm available, there are various weather stations located all over the country, reporting to The Norwegian Meteorological Institute. The institute makes this data available through its online Frost API[16]. By pairing every farm with their closest weather station, we got weather data that should be similar to the weather at the farm itself. As there are significantly more farms than weather stations, many farms share the same weather data. As there were 353 weather stations with precipitation data and 288 with temperature data, the accuracy is limited. However, the stations are still spread out so that the median distance between the farm and its closest precipitation weather station is 7.3 km, and the average is 9.0 km. As the weather data adds many data points, and because short term changes are unlikely to affect production, it has been aggregated by several different time spans to speed up training and reduce overfitting.

4.1.5 Combined data

Altogether, we end up with the following columns for our training data:

- latitude
- elevation
- area
- crop type
- estimated growth start date

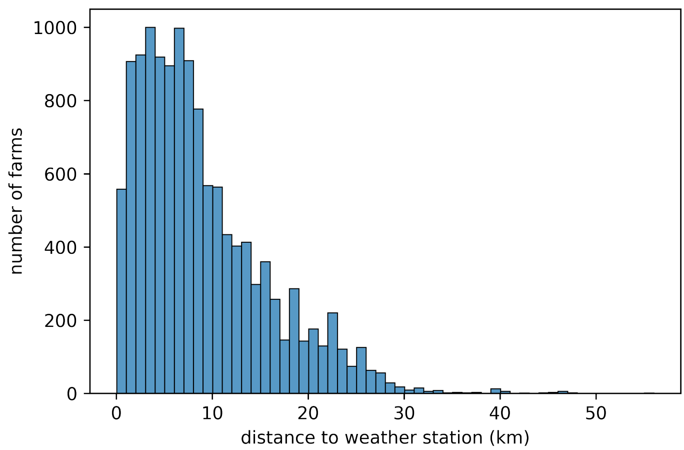


Figure 4.1: The distribution of distances to the farms' closest temperature sensor in 2019

Columns of daily aggregated weather data for the season:

- min temperature
- max temperature
- mean temperature
- precipitation (mm)

Columns for historical data (four previous years):

- barley (kg)
- wheat (kg)
- oat (kg)
- rye (kg)

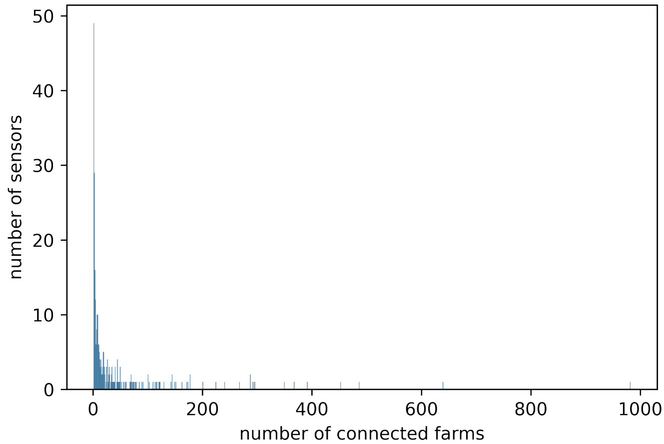


Figure 4.2: The distribution of farms per temperature sensor in 2019

4.2 Filtering data

Before feeding the neural networks with data, we perform some data analysis on the different datasets to better understand and process the data in a meaningful way.

4.2.1 Grain deliveries

The grain deliveries dataset contains entries for every farmer and every year. Together with the grants dataset, the grain deliveries allow us to look at each farmer's yield compared with how many hectares of a field is harvested. From figure 4.3, we see that yield is closely dependent on the farm's size (area harvested), yet there are still large variations on how much yield different farmers can get from a similar area harvested.

However, by looking closer at figure 4.3, there are apparent discrepancies in the reported raw data. A few reports include negative yield, and there are a few entries where a farmer has reported harvesting of many hectares but has a yield of zero.

We see most farmers manage a yield somewhere between 0 and 10 tonnes per hectare. Some farmers supposedly manage an extremely high yield, over 20 tonnes per hectare and higher. According to SSB [29], the average Norwegian farmer produced between 3 and 5 tonnes per hectare each season from 1997 to 2012, across the four major grain types. This average fits well with the majority of the samples in our dataset, and at the same time, indicates that the observed samples with very high yield could be based on invalid or incomplete reports.

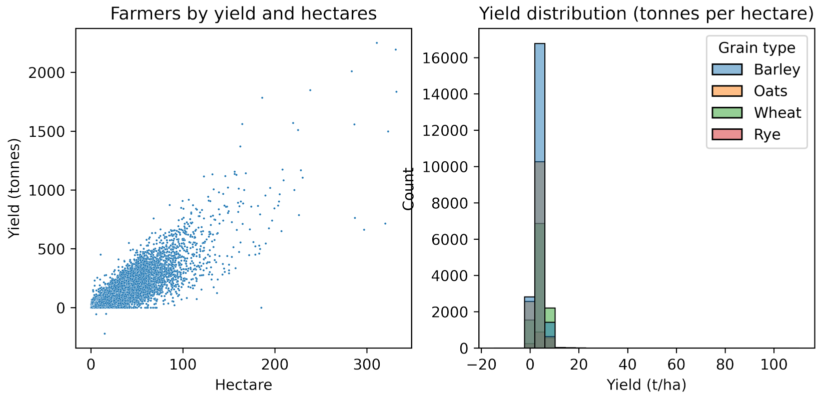


Figure 4.3: Each farmers yield and yield distribution (per grain type), before filtering.

Before using the data for training the neural network, we filter the data by removing outliers based on the distance from the mean yield. Because the different grain types' distributions are different, this filtering is done separately for each grain type. Removing data points more than two standard deviations from the average yield results in a distribution that closely resembles a normal distribution (see figure 4.4).

4.3 Data preparation

4.3.1 Normalization

All features are scaled or normalized before used as input to the neural network. Min-max normalization was applied to most features, meaning the values are linearly mapped to a value between 0 and 1, based on the

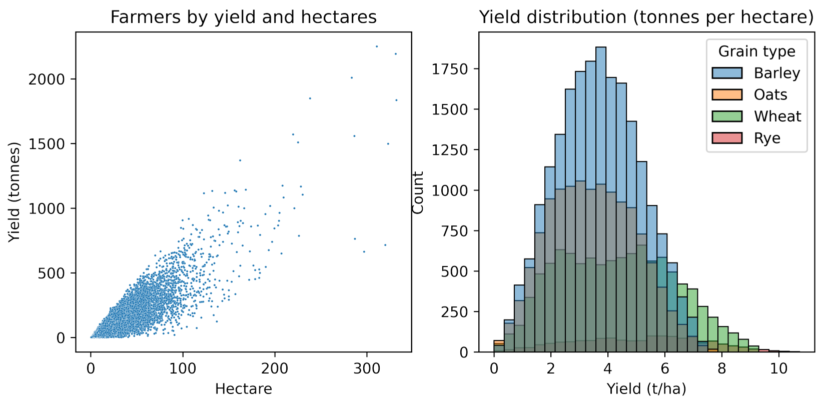


Figure 4.4: Each farmers yield and yield distribution (per grain type), after filtering.

maximum and minimum values of that feature.

$$normalized = \frac{value - minimum}{maximum - minimum}$$

For weather features, we normalize slightly differently. To keep temperature and precipitation values consistent across all the weather features, we apply a predefined constant value for the lower and upper normalization limits. The formula for normalization then becomes:

$$normalized = \frac{value - lower}{upper - lower}$$

We use a lower value of -30 degrees and an upper value of 30 degrees for temperatures, meaning that values in the range [-30, 30] are linearly scaled to [0, 1]. We use 0 and 10 mm as lower and upper values for precipitation, and for historical yield data, we use 0 and 10000.

4.3.2 Denormalization

Because the target value (y) is normalized, it makes the neural network's output values difficult to interpret compared to real-world yield numbers. In the following sections, we sometimes denormalize the output values using the following formula:

$$value = output \times (maximum - minimum) + minimum$$

4.4 Neural Network

Firstly, predicting yield is ultimately a regression task, and many valid techniques could be utilized to solve this. As discussed in section 3 State-of-the-Art, O'Neal et al. [23] saw that ANNs outperformed regression models for maize yield prediction in 2002, which indicates that neural networks also could be effective for grain yield prediction.

4.4.1 Network Structure

The model was created using Tensorflow and is a simple network built up by dense layers. The network consists of one input layer, three hidden dense layers with dropout in-between, and a final output layer.

The input layer consists of all 762 features available, fed into the first hidden dense layer of 256 neurons. Next, a dropout layer with a rate of 0.1 to add random noise before another hidden dense layer of 64 neurons, another dropout layer with a rate of 0.25 leading to the last hidden layer of 64 neurons. Last, there is an output layer of one single neuron with linear activation, which outputs the predicted value. See figure 4.5 for a graphical representation.

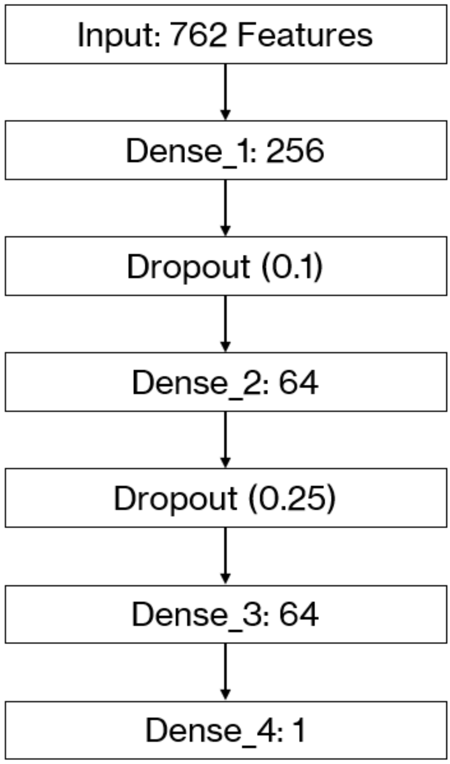


Figure 4.5: Structure of the neural network

After experimenting with the activation functions sigmoid, tanh, and relu in the hidden layers, we saw that tanh performed slightly better for this problem. The output layer did not have an activation (i.e., linear activation). When training the model, we saw the best result using the Adam optimizer with a learning rate of 0.0001 and a batch size of 512.

Part II

Experiments and Results

Chapter 5

Results

This chapter presents the results found by using the Neural Network to predict aspects of grain yield on farms throughout Norway.

5.1 Predicting total yield

Initial results from feeding the neural network with weather data, historical yield, grants information, and farm data (the size and type of crop harvested) while training it with each farmer's yield are presented in figure 5.1. It reaches a validation loss of 0.008 mean absolute error, which roughly translates to an average error of 18 tonnes per farmer [1].

[1] The network is fed with yield data normalized between 0.0 and 1.0, and thus outputs most values in this range as well. The highest yield in the dataset is 2250 tonnes, and thus an average (normalized) value of 0.008 equals 18 tonnes.

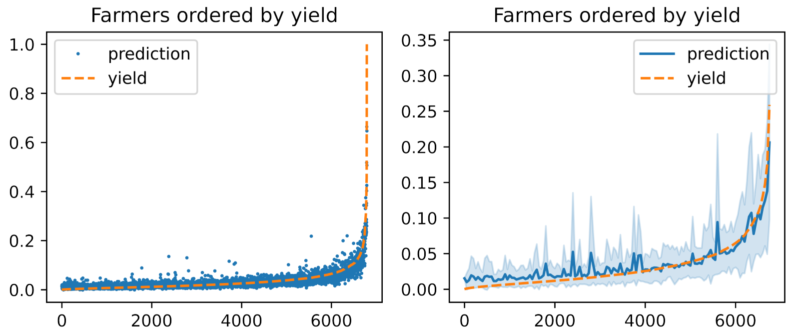


Figure 5.1: Neural Network predictions compared with the actual yield on validation data

Both plots show the data ordered by yield, in increasing order. The left plot shows each prediction as a point, while the right plot is a smoothed line plot that shows the median prediction and a minimum and maximum for bins of 50 samples each.

5.2 Predicting yield per hectare

As the harvested area increases, so does the weight of the total yield. To understand how good or bad a harvest is, the yield per hectare is a better measurement. Without changing any other input to the neural network, we set the target value to the yield per unit of planted area. The NN consistently managed a mean absolute error of 0.086 after training, which, when we denormalize, is roughly equal to 930 kg per hectare, 24% of the average yield per hectare in the dataset (3828 kg). For the average farmer in our dataset, this equals 16.3 tonnes total yield, which is slightly better than predicting the total yield.

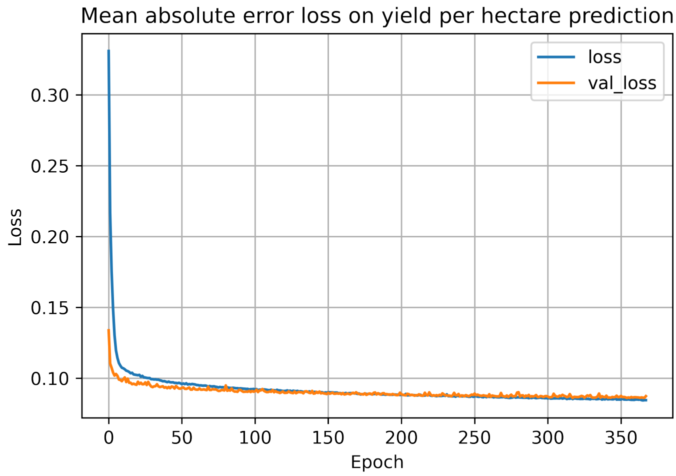


Figure 5.2: NN loss and validation loss during training on yield per hectare

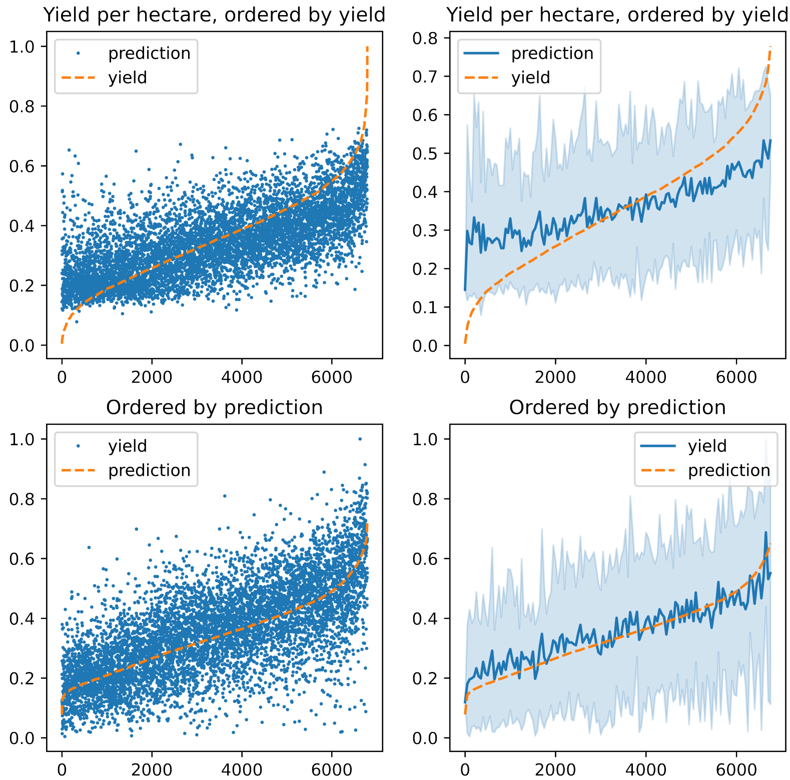


Figure 5.3: Neural Network predictions compared with the actual yield on validation data

Top plots show predictions ordered by actual yield per hectare, in increasing order, while the bottom plots show actual yield but ordered by predictions. The left plots show each prediction as a point, while the right plots are smoothed line plots that show the median prediction and a minimum and maximum for bins of 50 samples each.

Chapter 6

Experiments

6.1 Grants used as a proxy for area

There is data available for grain deliveries in Norway from 2013 to 2019 and each farmer's grants for these seasons. However, the public grants data changed from 2017 onwards to include more information such as area harvested per grain type. From 2013 to 2016, the grants included only one property of how much money the farmer got for the total area utilized: grants per area.

This experiment was to consider using the single feature of grants per area and compare the results to what was achieved in the section 5 Results. When there is a single feature representing the total area per farm, it is harder to distinguish between how much was grown per grain if the farmer delivered multiple types. Besides, the grants per area also include grass harvested areas and cases where the grain crops are harvested for animal fodder use (i.e., not delivering the produced grains).

Despite this, with the extra three years of grain delivery data, the model was able to train and predict reasonably well the farmer's relative production.

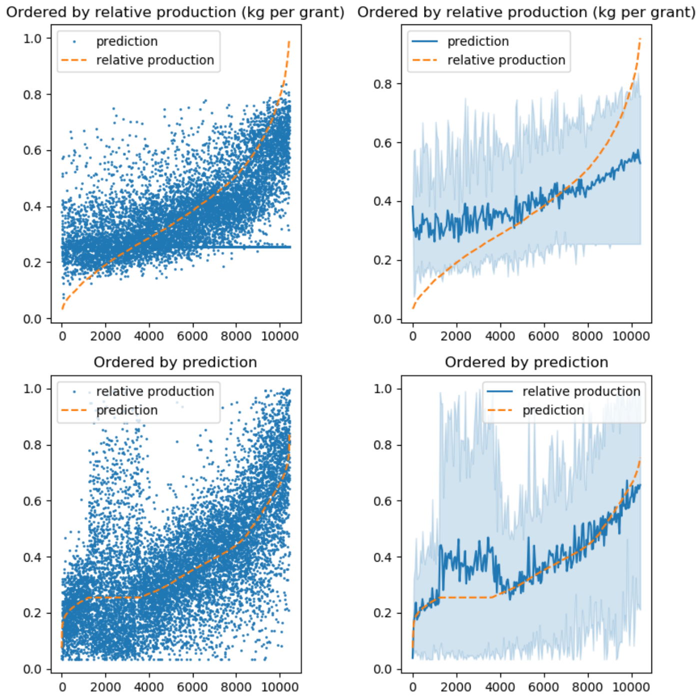


Figure 6.1: NN using grants as a proxy for area to utilize all available data.

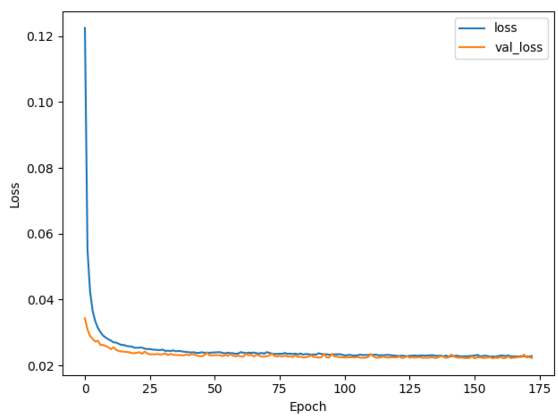


Figure 6.2: NN loss and validation loss during training on yield per relative production kg by grants

It is not easy to directly compare the results described in section 5.2, which used the area feature directly to the results in figure 6.1 using grants as area. Using grants gives more data, but it is less granular, and the amount of grants per hectare is fluctuating from year to year and from commune to commune. We see that the model reaches a loss of 0.115, which is 0.028 higher than what was achieved with the more detailed area in section 5.2.

We also saw a tendency of the model to guess the overall average, as shown in the bottom right plot of 6.1.

6.2 Predicted yield with different grain type

When we had created and trained a model, we did further experiments to see how alternative scenarios could affect the yield. For this purpose, we deployed the same model as proposed in 5 Results and comparing the results to what the model predicted if the yield were of another type, while factors such as weather, utilized area, etc., remain the same. This results in five data points for the same farmer, four alternative yields, and one for the actual yield. This may indicate whether another grain type could have granted a better yield or how the actual yield compared to the model's prediction of the same type. See figures 6.3 and 6.4.

In figure 6.3, we see a farmer who chose to plant barley. This farmer had slightly more yield per hectare than what the model predicted barley would produce, which indicates that the farmer performed well. We can also see that wheat or rye could have granted an even bigger yield had the farmer chosen any of these grains at the beginning of the season.

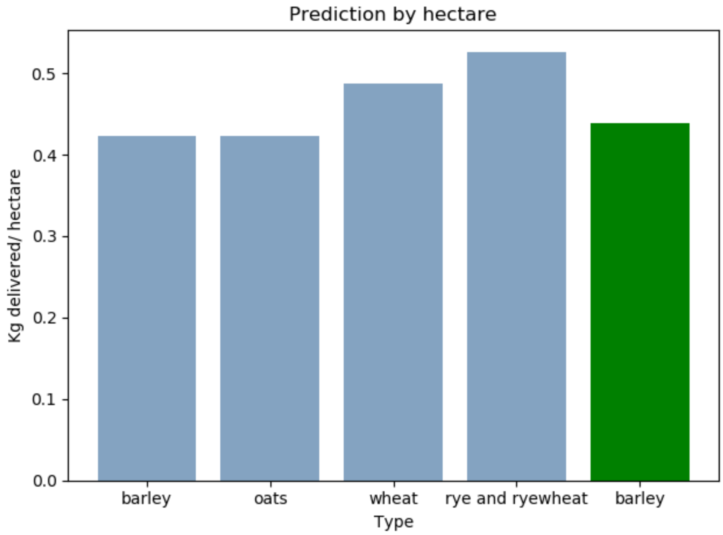


Figure 6.3: Different production of a farmer 1.

This farmer produced barley (green bar). Comparing the prediction and the actual yield could indicate that the farmer performed better than the models predict

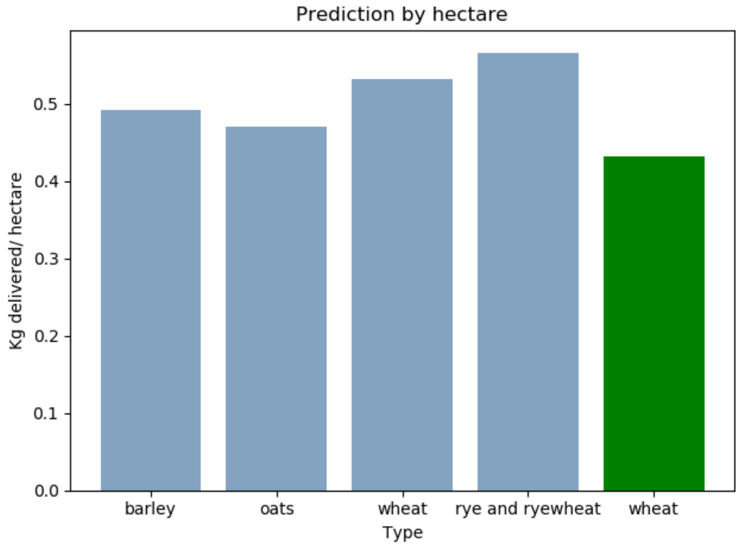


Figure 6.4: Different production of a farmer 2.

This farmer produced wheat (green bar). Comparing the prediction and the actual yield could indicate that the farmer performed slightly worse than anticipated.

6.3 Effects of weather on prediction

To test how well the neural network can generalize on weather data, we set up an experiment where we use data from a year with meager yield and apply weather data from a year outside the training data with a much better average yield. The dataset from 2018 was chosen as the base dataset, as this was the year that saw the lowest total grain production since 2013[22][27][28]. We then applied weather data taken from 2015, which is not included in any training data and is the year with the highest grain production since 2013 [22][27][28].

Both actual 2018 data and data with weather data replaced with data from 2015 were fed to the model, and the results (see figure 6.5) show that the network can accurately predict (on average) the 2018 data. The network also predicted a 100 kg per hectare increase (34% increase) with weather data from 2015, consistent with that year's much higher grain production.

However, the actual average yield from 2015 was even higher than the predicted 377 kg (denormalized) per hectare [28].

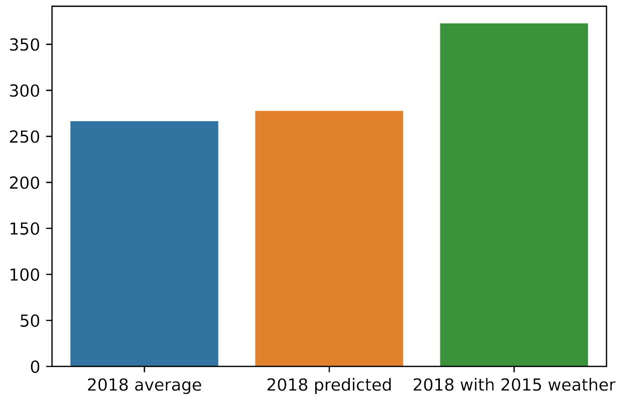


Figure 6.5: The effects of applying weather data from a year with known good weather.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

The challenges of accurate yield prediction are many. While weather and climate factors are a major factor determining the yield, it does not fully explain observed differences in farmers' yield.

By using a deep neural network with three hidden layers, yield prediction reaches an accuracy of 930 kg per hectare, using publicly available data from Norwegian grain farmers along with geographic location data, daily weather temperatures, precipitation, and historical production numbers. The neural network appears to generalize well on weather data, even when tested on data from an earlier year that was never included during training. By changing the input manually, the network can predict how much a farmer could have produced had they grown another type of grain, indicating that it could be possible to use similar models to help farmers decide what types of grain to plant.

7.2 Future Work

7.2.1 Additional Data Sources

Although data was acquired and processed at several stages throughout the project period, obtaining more data could improve learning. Some potential data sources were available but not included in this project. Others became available late in the project period or are expected to become available after this project.

Sentinel-2 Satellite Imagery

The Sentinel-2 Satellites provide satellite images of the earth, usually with 5-10 days intervals, depending on the visibility conditions[24]. This could provide the model with additional input in ways described below and in 7.2.2.

Digifarm Field Polygons

Digifarm is a Norwegian company specializing in the application of machine learning in agriculture[6]. They have processed Sentinel-2 satellite images and extracted geographical polygons for individual fields. With this data, a better estimate for the area used for crops could be made. In combination with the Satellite Imagery described in 7.2.1, it could be used to monitor the crops' condition during the growth period.

More accurate weather data

While the Norwegian Meteorological Institute's data has frequent reports with several metrics, the distance between sensors and farms means that some farms get imprecise weather data. Some methods could provide geographically closer data. The most obvious one is to use more sensors. Netatmo is a company that sells consumer smart home products[15], including weather sensors. They also provide a public API, where it is possible to get measurements[25]. However, the API limitations in terms of request frequency and availability of historical data are uncertain. Given that they are user-installed, the accuracy of the sensors is also uncertain.

Management data

Getting additional input from the farmers regarding the cultivation of their crops could provide better insight into how these factors affect yield. This could be data like watering, pesticide use, sow date, harvest date, and so on. Given a good learning model, it should improve performance, as these are all factors that should affect yield to a significant extent.

7.2.2 Different Models

While there were many significant improvements from increasing the data sources and improving preprocessing, testing a greater variety of models will give more references to the relative performance of the current DNN model and potentially outperform it. Using the satellite imagery described in 7.2.1, a CNN could likely extract additional information that our current model cannot, potentially leading to better predictions. Likely candidates include when plants start sprouting, rate of growth, and harvest date.

References

[1] Andrea Sofie Aasvang. Nationen - Felleskjøpet får støtte til kornsatsing. https://www.nationen.no/landbruk/felleskjopet-far-stotte-til-kornsatsing/.

[2] Norwegian Agriculture Agency. Produksjons- og avløsertilskudd til jordbruksforetak søknadsomgang 2019. https://data.norge.no/datasets/e341762d-311b-46f1-be25-078570f63204.

[3] Norwegian Digital Learning Arena. Bygg. https://ndla.no/subjects/subject:41/topic:1:188696/topic:1:189185/resource:1:82133. Accessed: 2020-12-01.

[4] Norwegian Digital Learning Arena. Havre. https://ndla.no/subjects/subject:41/topic:1:188696/topic:1:189185/resource:1:82136?filters=urn:filter:b5156afd-e11c-4426-a3ba-a1e25c03f489. Accessed: 2020-12-01.

[5] Norwegian Digital Learning Arena. Rug. https://ndla.no/nb/subjects/subject:41/topic:1:188696/topic:1:189185/resource:1:82142?filters=urn:filter:b5156afd-e11c-4426-a3ba-a1e25c03f489. Accessed: 2020-12-01.

[6] Digifarm AS. From crop to cloud. https://digifarm.io/.

[7] Govinda Bhandari. Effect of rainfall on the yield of major cereals in darchula district of nepal. *International Journal of Environment*, 3, 02 2014.

[8] Bondelaget. Norwegian agriculture - the norwegian model for agriculture and agricultural policy. https://www.bondelaget.no/getfile.php/13894650-1550654949/MMA/Bilder%20NB/Illustrasjoner/Norwegian%20Agriculture%20EN.pdf.

[9] D. G. Bullock. Crop rotation. *Critical Reviews in Plant Sciences*, 11(4):309–326, 1992. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07352689209382349.

[10] Encyclopaedia Britannica, rye. https://www.britannica.com/plant/rye. Accessed: 2020-12-01.

[11] Kolbjørn Eriksen. Klima og planteproduksjon i nord. https://nordnorge.nlr.no/publikasjoner/grovforskolen/1-klima-og-planteproduksjon-i-nord/, January 2018.

[12] Felleskjøpet. Norwegian agricultural purchasing and marketing Co-Op. http://www.fk.no/om-norske-felleskjoep/summary-in-english.

[13] Grainsa and RhinoReloaded. Why crop rotation is important in wheat production. https://www.grainsa.co.za/why-crop-rotation-is-important-in-wheat-production/.

[14] Ann Katrin Holtekjølen, Anne Uhlen, Einar Kr Holtet, and Åshild Ergon. hvete. http://snl.no/hvete, October 2019.

[15] Crunchbase Inc. Netatmo - crunchbase company profile & funding. https://www.crunchbase.com/organization/netatmo/. Accessed: 2020-12-12.

[16] The Norwegian Meteorological Institute. Frost API. https://frost.met.no/index.html, November 2020.

[17] Bonnie F. Jacobs, John D. Kingston, and Louis L. Jacobs. The origin of grass-dominated ecosystems. *Annals of the Missouri Botanical Garden*, 86(2):590–643, 1999.

[18] Brøndbo Karstein. Havre-NM 2019 - FoU - Alt om norsk havre. http://norskhavre.no/alt-om-norsk-havre/fou/havre-nm-2019.

[19] Saeed Khaki, Lizhi Wang, and Sotirios V. Archontoulis. A CNN-RNN Framework for Crop Yield Prediction. *Frontiers in Plant Science*, 10, 2020. Publisher: Frontiers.

[20] Kjersti Kildahl. Nine facts about norwegian agriculture - nibio. https://www.nibio.no/en/news/nine-facts-about-norwegian-agriculture.

[21] J. Liu, C. E. Goering, and L. Tian. A NEURAL NETWORK FOR SETTING TARGET CORN YIELDS. *Transactions of the ASAE*, 44(3):705, 2001. Place: St. Joseph, MI Publisher: ASAE.

[22] Brød og Korn. Kornproduksjon i Norge. https://brodogkorn.no/fakta/kornproduksjon-i-norge/.

[23] Monte R. O'Neal, Bernard A. Engel, Daniel R. Ess, and Jane R. Frankenberger. AE—Automation and Emerging Technologies: Neural Network Prediction of Maize Yield using Alternative Data Coding Algorithms. *Biosystems Engineering*, 83(1):31–45, September 2002.

[24] Sentinel Online. Sentinel-2 - Missions - Sentinel Online. https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2.

[25] Netatmo SAS. Weather api documentation. https://dev.netatmo.com/apidocumentation/weather.

[26] A. O. (nlvf Skjelvaag. Temperature map compiled by last squares interpolation. *Norwegian Agricultural Research*, 1987.

[27] SSB. Jordbruk. https://www.ssb.no/jord-skog-jakt-og-fiskeri/faktaside/jordbruk. Accessed: 2020-11-23.

[28] SSB. Korn og oljevekster, areal og avlinger, 2018/2019, foreløpige tall. https://www.ssb.no/jord-skog-jakt-og-fiskeri/statistikker/korn/aar-forelopige/2020-01-28.

[29] SSB. Tabell - Totalavling og avling i kilo per dekar av ulike kornslag. Fylke. https://www.ssb.no/150124/totalavling-og-avling-i-kilo-per-dekar-av-ulike-kornslag.fylke. Accessed: 2020-11-24.

[30] Tine. Family farming: the key to food production in norway. https://www.tine.no/english/about-tine/family-farming-the-key-to-food-production-in-norway. Accessed: 2020-11-23.

[31] Oregon State University. Environmental Factors Affecting Plant Growth. https://extension.oregonstate.edu/gardening/techniques/environmental-factors-affecting-plant-growth. Accessed: 2020-12-07.

[32] Thomas van Klompenburg, Ayalew Kassahun, and Cagatay Catal. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177:105709, 2020.

[33] Mauritz Åssveen and Unni Abrahamsen. Varmesum for sorter og arter av korn. *Grønn Forkning*, 2(99):55–59, 1999.

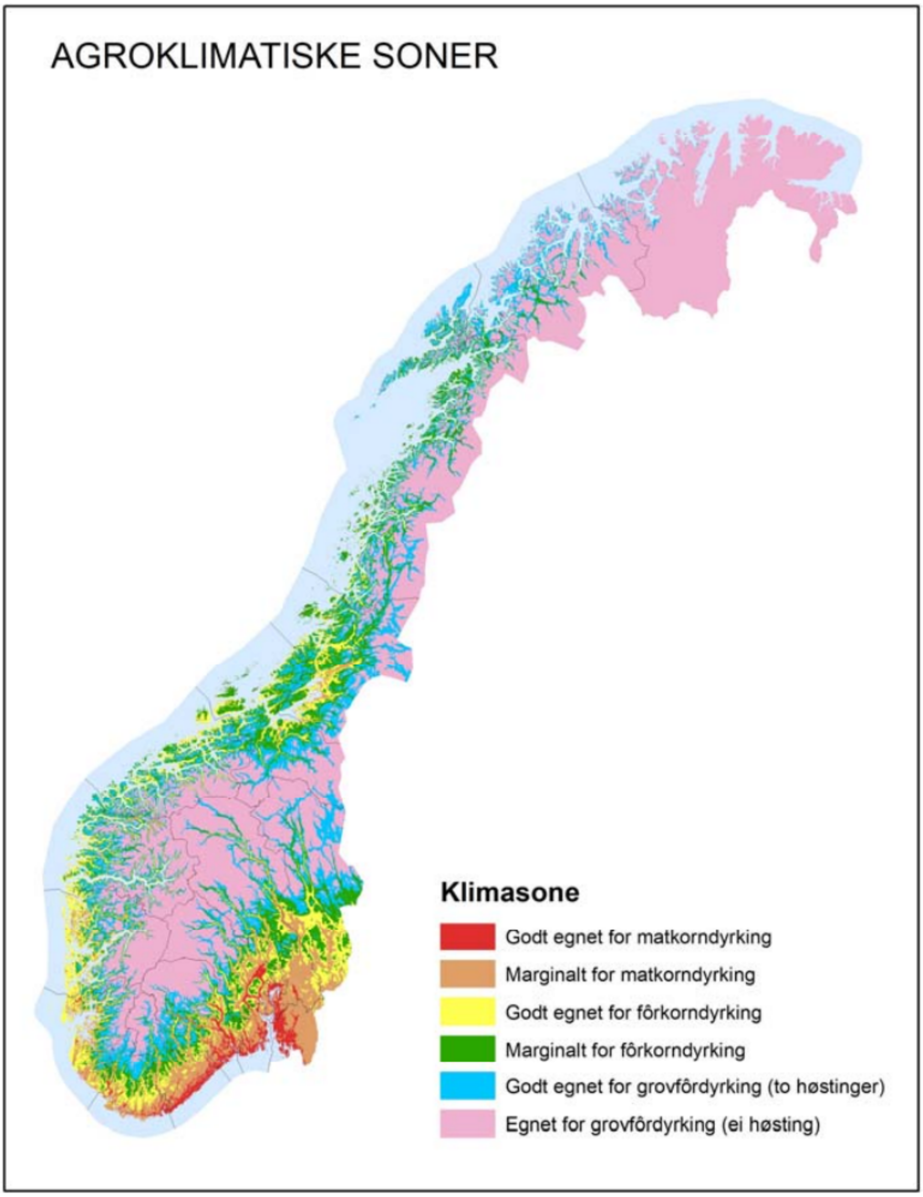
[34] Rozman Črtomir, Cvelbar Urška, Tojnko Stanislav, Stajnko Denis, Pažek Karmen, Martin Pavlovič, and Vračko Marjan. Application of Neural Networks and Image Visualization for Early Forecast of Apple Yield. *Erwerbs-Obstbau*, 54(2):69–76, June 2012.

Appendices

A Project code

https://github.com/putetrekk/kornmo

B Graphical distribution of agro-climatic zones in Norway



UiA University of Agder
Master's thesis
Faculty of Engineering and Science
Department of ICT