

Article

Identification of Optimal Starting Time Instance to Forecast Net Blotch Density in Spring Barley with Meteorological Data in Finland

Outi Ruusunen ^{1,*}, Marja Jalli ², Lauri Jauhiainen ², Mika Ruusunen ¹ and Kauko Leiviskä ¹

¹ Control Engineering Research Group in Environmental and Chemical Engineering Research Unit, Faculty of Technology, University of Oulu, FI-90014 Oulu, Finland

² Natural Resources Institute Finland, Tietotie 4, 31600 Jokioinen, Finland

* Correspondence: outi.ruusunen@oulu.fi

Abstract: The performance of meteorological data-based methods to forecast plant diseases strongly depends on temporal weather information. In this paper, a data analysis procedure is presented for finding the optimal starting time for forecasting net blotch density in spring barley based on meteorological data. For this purpose, changes in the information content of typically measured weather variables were systemically quantified in sliding time windows and with additionally generated mathematical transformations, namely with features. Signal-to-noise statistics were applied in a novel way as a metric for identifying the optimal starting time instance and the most important features to successfully distinguish between two net blotch densities during springtime itself. According to the results, the information content of meteorological data used in classifying between nine years with and four years without net blotch reached its maximum in Finnish weather conditions on the 41st day from the beginning of the growing season. Specifically, utilising weather data at 41–55 days from the beginning of the growing season maximises successful forecasting potential of net blotch density. It also seems that this time instance enables a linear classification task with a selected feature subset, since the averages of the metrics in two data groups differ statistically with a minimum 68% confidence level for nine days in a 14-day time window.

Keywords: advanced data analysis; feature generation; plant disease prediction; signal-to-noise statistics; modern agriculture



Citation: Ruusunen, O.; Jalli, M.; Jauhiainen, L.; Ruusunen, M.; Leiviskä, K. Identification of Optimal Starting Time Instance to Forecast Net Blotch Density in Spring Barley with Meteorological Data in Finland. *Agriculture* **2022**, *12*, 1939. <https://doi.org/10.3390/agriculture12111939>

Academic Editors: Dimitrios Moshou and Xanthoula Eirini Pantazi

Received: 3 October 2022

Accepted: 15 November 2022

Published: 17 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The performance of the modern data-based forecasting tools for plant diseases is highly dependent on the methods applied and on the representativeness of the available data. The temporal characteristics of the information content in meteorological data and its effect on the classification potential of net blotch risk levels have recently been discovered [1]. In this paper, the optimal starting time instance for net blotch risk forecasting in Finnish weather conditions is studied with an analysis framework including signal-to-noise statistics for meteorological data with feature generation proceeding in sliding data windows.

Barley, *Hordeum vulgare* L., is one of the largest grain crops and in 2020 it was grown on 51.6 million hectares globally [2]. Barley is grown for example as animal fodder and as a source of malt for beverages. It is also common in food products such as breads, soups and stews and in health products. There are several biotic and abiotic pressure factors that challenge barley production. Barley net blotch is one of the most common fungal diseases in barley and is caused by the two following pathogens: *Pyrenophora teres f. teres* (net form) and *Pyrenophora teres f. maculata* (spot form). In Finland, net blotch was present in 86% of barley fields investigated in 2009 [3]. According to Jalli et al. [4], leaf blotch diseases, with a severity >50% at DC 73–77, cause an average of 1114 kg ha⁻¹ yield-loss in spring barley in the long term in the Nordic countries. An assessment in 2015 showed that 40% of fields in

South Tigray, in Ethiopia, had net blotch and 60% of them had its relative, spot blotch [5]. This can decrease the barley crop by 10–20% of the annual average yield [6,7], but yield losses as high as 40% have been reported [8].

There are several means to combat net blotch and also other foliar diseases: using clean seed from a resistant cultivar, utilising crop rotation and controlling nitrogen. Usually, chemical and biochemical means are additionally needed [9]. Chemical protection saves the crop, but the overuse of pesticides should be avoided. In [10], the benefits and hazards of pesticides are discussed with many examples. The authors look at pesticide use from different aspects, namely the exposure to pesticides of production workers, formulators, sprayers, mixers, loaders and farm workers. Furthermore, the impact of pesticide residues through food commodities are examined widely [10]. The authors summarise this as “Pesticides have contaminated almost every part of our environment” and advocate for finding ways to protect people against the adverse effects of pesticides. There must be a balance between chemical crop protection and the risks caused by pesticides. In the European Union, IPM (Integrated Pest Management) is codified into the form of a directive which needs to be followed by farmers. According to the directive, chemical protection needs to be justified and well-documented [11]. The main idea is to avoid the negative impacts of agrochemicals and use chemical protection only when absolutely necessary. Another driving force is the fact that the European Commission has adopted the proposal to restore damaged ecosystems and nature by 2050 and to halve the use of pesticides by 2030 [12]. This strongly dictates the reduction of pesticides and will lead to more sustainable food systems in the future.

Forecasting is an important tool for the early detection of plant diseases and in evaluating the risks connected to them. It can help in choosing and implementing disease management strategies. The increased amount of information and improved possibilities to process it has made forecasting tools viable for everyday use. A couple of reviews have shown increased interest in these applications [13–15].

In practice, the aim of forecasting applications is to avoid routine pesticide sprayings and to help farmers in decision making when planning their chemical crop protection strategy. Risk assessment is often based on pathogen- and plant-specific factors, selected weather parameters, agronomic variables e.g., cultivar resistance or disease pressure, and in some cases on earlier infection data and geological location. Three different risk models and their use in several test fields in five different countries in the Nordic–Baltic region were studied in [16]. In their paper, the models discussed were the Danish decision support Crop Protection Online (CPO), the Danish Humidity Model (HM) and the Finnish net blotch and scald model WisuEnnuste. The authors compared the models’ suitability to predict barley leaf blotch diseases. In the CPO system, the risk assessment for all relevant barley diseases is computed by the number of days with precipitation over 1 mm, information about cultivar resistance and disease data [17,18]. The Danish Humidity Model originally estimated the risk for *Septoria tritici* blotch in winter wheat and is based on rain events such as the relative humidity or leaf wetness [19]. The Finnish WisuEnnuste [20] has been developed to estimate the field-specific disease risk based on information about the previous crop, the tillage method, the cultivar resistance and certain weather parameters. In [21] the Fourier transform is used in studying the effects of intra-day meteorological changes to *Septoria* net blotch in winter wheat.

Some examples of plant disease prediction tools for decision making and crop protection are documented and discussed in [22]. The authors focused on *Fusarium* head blight, which is the major fungal disease that causes losses in wheat and barley production in Canada. The use of fungicides in addition to evolutionary factors have led to more virulent forms of *Fusarium* head blight. Fernando et al. discuss the utilization of modern prediction tools for plant diseases as well as potential plant defence mechanisms and resistance breeding as a means for plant disease management.

Some other prediction systems for *Fusarium* head blight are presented in [23] and [24]. One web-based platform that allows *Fusarium* risk assessment based on parameters such

as the geographic location, crop type and weather is the Fusarium Head Blight Prediction Centre in the United States [25–28]. Generally, the need for an accurate forecasting system for crop protection has been recognized and several applications have been developed around the world.

In [29], a prediction system for barley net blotch is presented and discussed with detail. The computation of the risk index for barley net blotch utilises selected weather variables in different growing zones in Finland. Instead of using original variable values in forecasting, the forecasting accuracy was increased by using features generated from the original data. The feature selection utilised the two-sample *t*-test. The data originated from the open weather data of the Finnish Meteorological Institute and long-term observations of plant disease severity in different growing zones in Finland; forecasting was performed without field-specific measurements. In that study, the forecasting of barley net blotch densities was carried out with advanced data fusion applied to two different data sets.

The accuracy of data-based forecasting depends on the method applied and on the information content of the utilized data. In [1], it is demonstrated that the amount of information content in data is time-dependent. This means that the accuracy of plant disease forecasting may vary during the growing season. Three different data window sizes (7, 14 and 21 days) were studied in the paper, while the starting point of the prediction varied between zero and 50 days from the beginning of the growing season.

The previously mentioned two papers [1,29] show that feature generation improves the forecasting accuracy and helps to avoid additional field tests and the forecasting accuracy depends on choosing the correct time sample, especially the starting point of forecasting. The research problem in this paper is how to identify and define the optimal starting time instance for the net blotch forecasting in Finnish conditions, so enhancing the performance of plant disease forecasting methods that utilise meteorological data. Based on the earlier results [1], this study uses the window size of 14 days in forecasting. The target is to be able to define automatically the optimal starting point for forecasting from the history database by evaluating the information content of data for every time-step from the beginning of the growing season. The signal-to-noise ratio was used as the metric for the information content of the data.

2. Materials and Methods

2.1. Data

Weather data from the open database of the Finnish Meteorological Institute (FMI) and field observations of net blotch density from the official variety trials database of the Natural Resources Institute Finland (Luke) were utilised. The net blotch data was collected during the period 1991–2017 and the test fields were located in Central and Southern Finland. Net blotch density is divided into two categories:

- Category 1 (very low net blotch density, maximum net blotch severity value of 0.5%);
- Category 2 (net blotch appears in the selected observation fields in these years, severity value of 0.6–5%).

One example of labelling the intensity of plant disease in cereals is presented in [30]. The locations of the test fields and the years of the selected weather data by category are presented in Table 1.

A description of the weather data and pre-processing is presented in detail in [29]. Information on the weather stations related to the data used can be found in Appendix B. In this study, the weather variables analysed were:

- Atmospheric pressure (hPa);
- Relative humidity (RH %);
- Temperature (°C);
- Dew point temperature (°C).

In the data analysis, the daily minimum, daily maximum and daily average values of the above variables were considered.

Table 1. Location of test fields and the years of utilised weather data by category.

Location of Test Fields	Mynämäki N = 6,732,402.033 E = 218,702.907	Jokioinen N = 6,746,822.331 E = 308,359.757	Seinäjäoki N = 6,986,750.229 E = 271,138.563	Siikajoki N = 7,174,584.799 E = 408,818.353	Years in Total
Years of observations Category 1	2011	2013	2011	2010	4
Years of observations Category 2	2013, 2014, 2016	2014, 2015	2016	2012, 2014, 2015	9

2.2. General Structure of Data Analysis

The weather variables were first normalised with linear scaling between 0 and 1 based on Equation (1):

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where the minimum value (x_{min}) and the maximum value (x_{max}) were found from the Category 1 data (low net blotch density). The Category 2 (high net blotch density) data were normalised with their corresponding minimum and maximum values. The beginning of the growing season varies according to the year and observation field because of the varying climate conditions and the geological position. This has been considered in the analysis by selecting the starting point of each data set at the beginning of the growing season instead of a fixed calendar date, as explained below in the Section 2.3.

The main idea of the analysis is to search and rank the time windows where yearly weather data is grouped within the categories, but to at the same time separate Categories 1 and 2. The general concept for analysing the classification properties of net blotch densities based on temporal weather conditions is presented in Figure 1. In addition to the original weather measurements, the feature prototypes (Appendix A), namely mathematical transformations of the original variables, were incorporated into the analysis.

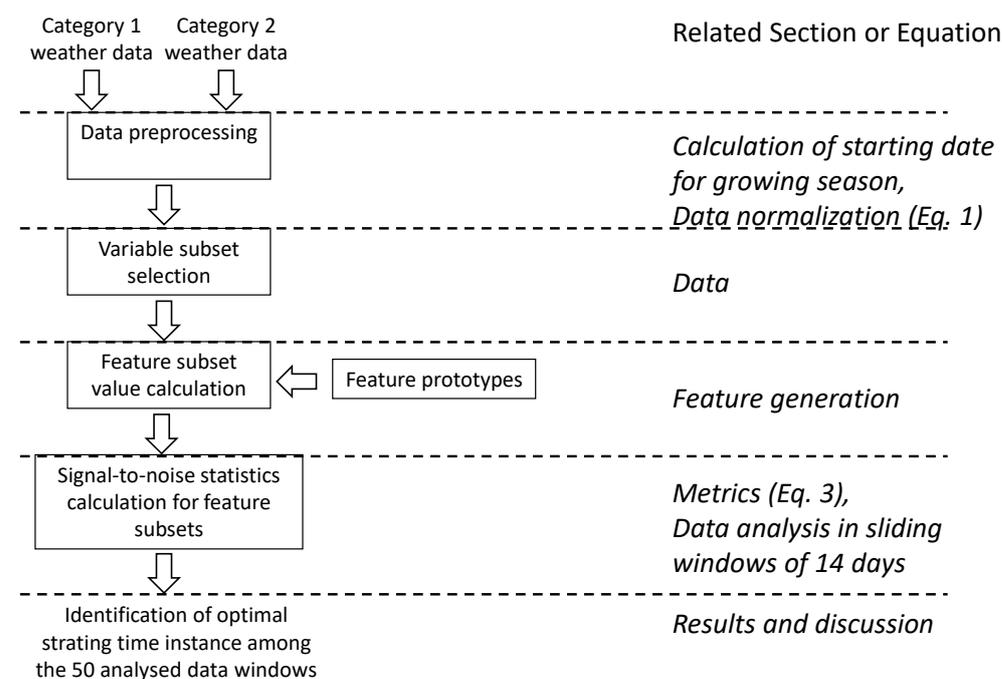


Figure 1. Analysis procedure for meteorological data to identify time instance of optimal information content concerning forecasting the severity of net blotch occurrence.

The feature subset values were calculated using the feature prototypes (see Section 2.4) at each time-step from the beginning of the growing season for fifty days onwards, proceeding in sliding windows of fourteen days. The feature values were thus the sum of daily signal-to-noise ratios over the sliding window (Equation (2)). After all the time-steps and feature combinations had been calculated, the optimal starting point was determined as the time instance related to the highest obtained feature values. The 14-day window for data analysis was then repeated 50 times, with the first time window starting from the beginning of the growing season. The recurrence of data windows in the analysis is illustrated in Figure 2.

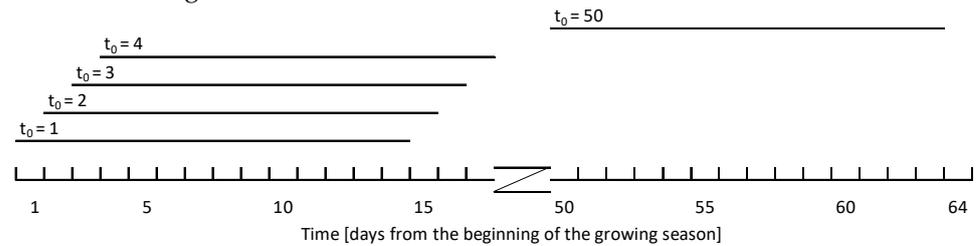


Figure 2. Recurrence of analysis in sliding time windows.

2.3. Starting Date of Growing Season, Automatic Calculation

In this research, the starting date of the growing season is calculated using weather measurements. The beginning of the growing season is determined as the time when the mean outdoor temperature remains over $+5\text{ }^{\circ}\text{C}$ for 10 consecutive days. The estimated time for the beginning of the growing season instead of a certain calendar date, for example the sowing date, enables here the spatially-independent comparison of data sets that may also exhibit different weather conditions related to the measurement location. This especially results in a standard and automatic procedure for triggering the data analysis. This method thus differs considerably from the typical usage of the sowing date, requiring the manual and field-specific insertion of the date into present disease prediction tools.

The data sets applied in this study exhibited differences between the sowing date and the beginning of the growing season. The actual sowing date is on the y-axis at zero value, and the year- and field-related difference (in days) from the beginning of the growing season is represented with the bars in Figure 3.

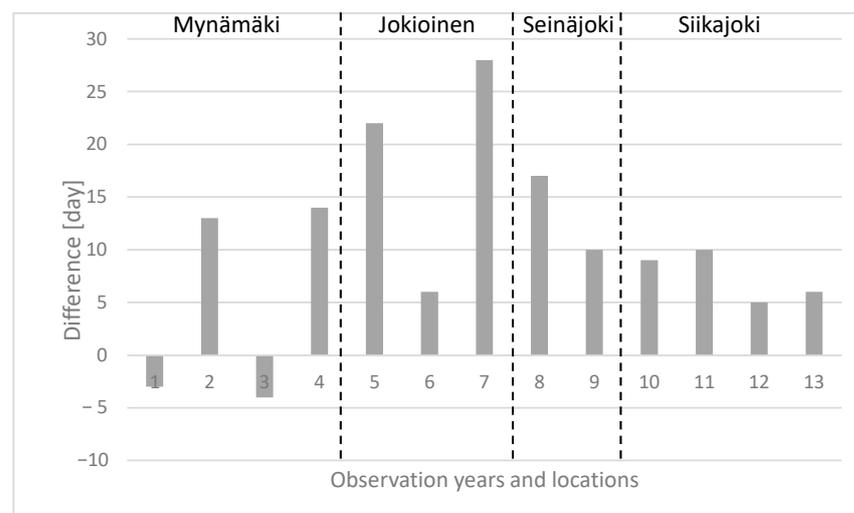


Figure 3. Difference between the sowing date and the beginning of the growing season, presented by year (running number on x-axis) and observation fields (names).

2.4. Feature Generation

Feature generation was utilised to extract more information for classification between net blotch densities than was available with the original weather variables [29]. Some examples of feature generation techniques are presented in [31–35]. In this study, the feature generation method presented in [36] (p. 50) was applied, and the considered feature prototypes are listed in Appendix A.

The feature subset value of every tested feature was computed with every possible variable combination in every fifty selected time-steps. The classification ability of the generated features was studied with signal-to-noise statistics (see Section 2.5). The generated features were normalised before applying the signal-to-noise statistics to ensure the comparability of the D_{sn} value (see Equations (2) and (3) below).

The total number of tested feature prototypes was 115 in 715 different combinations in groups of four variables, including generated features from a single to three variables. The number of generated features tested in each time window was 1,973,400. The features were generated as combinations of the minimum, maximum and average of the available weather variables (4) and the calculated Leaf Wetness Duration (LWD) that was computed here on an hourly basis as presented in [37] with rules and their inference as follows:

- If the relative humidity is >87%, then the leaf is humid → LWD = 1;
- If the relative humidity is >70%–<87% and increasing >3% per 30 min, then the leaf is humid → LWD = 1;
- If the relative humidity is >70%–<87% and decreasing >2% per 30 min, then the leaf is dry → LWD = 0;
- If the relative humidity <70%, then the leaf is dry → LWD = 0.

The calculated daily LWD was thus the sum of 24-hourly estimates according to the rule inference above.

2.5. Metrics

In this study, the signal-to-noise statistics were examined for the vectors (here time-series) applied for the classification of the different weather data sets according to the net blotch severity. For example, in [38] the authors have successfully utilised signal-to-noise statistics in the prediction of embryonal tumour outcomes in the central nervous system based on gene expression. The authors developed a classification system based on DNA microarray gene expression data and predicted the risk of selected tumour outcomes.

For signal-to-noise statistics, D_{sn} , the distance between two vectors a and b , Category 1 and 2, respectively, with their mean values μ_a and μ_b and standard deviations δ_a and δ_b are computed according to Equation (2) [38]:

$$D_{sn} = \frac{(\mu_a - \mu_b)}{(\delta_a + \delta_b)} \quad (2)$$

Equation (2) is applied in this study. The calculation of D_{sn} for the identification of the optimal starting time instance for net blotch prediction proceeds in sliding windows from the beginning of the estimated growing season and the following 50 days, step by step in data windows of 14 days for every generated feature n as a sum of the calculated D_{sn} daily values:

$$D_{sn} = \sum_{j=1}^{14} \frac{|\bar{x}_{nj}[\text{MC}_1(1), \text{MC}_1(4)] - \bar{x}_{nj}[\text{MC}_2(1), \text{MC}_2(9)]|}{s_{nj}[\text{MC}_1(1), \text{MC}_1(4)] + s_{nj}[\text{MC}_2(1), \text{MC}_2(9)]} \quad (3)$$

where MC_1 and MC_2 are feature matrices generated from scalar observations of weather variables related to the years with data of Categories 1 and 2 (see Table 1), \bar{x}_{nj} is the average of the data of the years in question and s_{nj} is the standard deviation of the same data. In Figure 4, the behaviour of the D_{sn} index is illustrated. There, Feature number 1 would be ranked as a more plausible candidate than Feature number 2 by comparing their

calculated D_{sn} values in classification of the two groups (o and x). According to Figure 4 and Equation (2) with the same notation of statistical quantities, the resulting value of D_{sn} for Feature 1 would be much higher than for Feature 2 in this case.

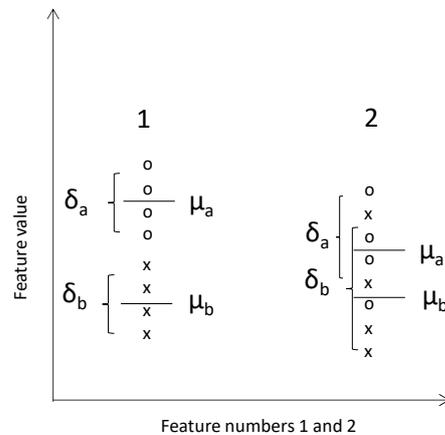


Figure 4. An illustrated example of the usage of D_{sn} for identifying classification properties of features in the case of two data groups (o and x). Classification between the data groups would be successful with Feature 1 in this case.

3. Results and Discussion

In the following figures, the resulting D_{sn} values in the fourteen-day time window are presented for each time step. The daily average, minimum and maximum of the studied variables were tested and the D_{sn} values were calculated accordingly to every analysed feature subset (Equation (3)).

In Figure 5, the calculated D_{sn} values for the average, minimum and maximum outdoor temperature are presented. The highest D_{sn} value, 52.5, is achieved on day 40 from the beginning of the growing season with the daily average outdoor temperature. The D_{sn} value in the case of the average outdoor temperature remains relatively high between days 37 and 40, but then the value falls rapidly to 28.2 on day 41. On the other hand, the D_{sn} value of the maximum outdoor temperature achieves its highest value, 43.3, on the same day while the highest D_{sn} value of the minimum outdoor temperature of 29.3 is achieved on day 29. The D_{sn} value of the average outdoor temperature higher than the D_{sn} values of the minimum and the maximum outdoor temperature during the studied period. This indicates that the information content of the outdoor temperature related to the appearance of barley net blotch is the highest during the two weeks starting on days 36–40 from the beginning of the growing season.

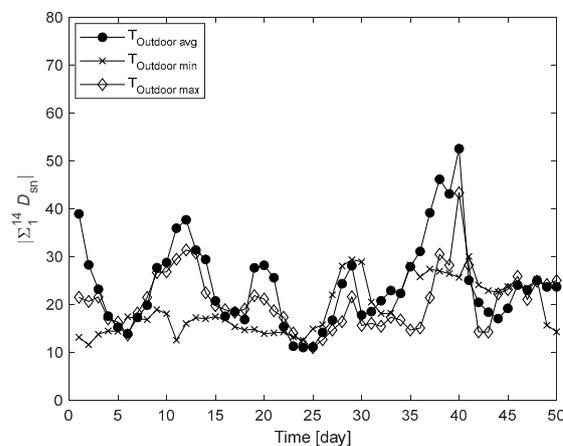


Figure 5. Variation of the classification potential (D_{sn} values) with daily minimum, maximum and average outdoor temperatures. Day 1 on the x-axis is the first day of the growing season.

In Figure 6, the D_{sn} values of the average, minimum and maximum daily relative humidity are presented. The highest D_{sn} value, namely 57, is achieved on day 22 from the beginning of the growing season with the minimum daily relative humidity. The average of the daily relative humidity reaches its highest D_{sn} value of 45.9 at almost the same time, namely on day 25. The maximum daily relative humidity exhibits the highest D_{sn} values in the time window 42–49 days from the beginning of the growing season. As can be seen in Figure 5, the D_{sn} values are relatively high between days 17 and 30 with all three variables, but the maximum daily relative humidity does not achieve its highest value of 49.5 until day 49.

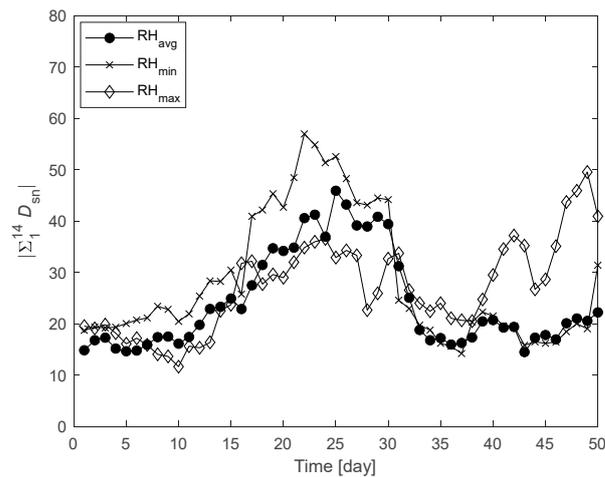


Figure 6. Variation of the classification potential (D_{sn} values) when using the daily minimum, maximum and average values of relative humidity. Day 1 is the beginning of the growing season.

In Figure 7, the D_{sn} values of the daily average, minimum and maximum dew point temperature are presented. The highest D_{sn} value, 48.9, is achieved for this weather variable on day 39 from the beginning of the growing season (day 1) with the daily average dew point temperature. In the case of the daily maximum dew point temperature, the D_{sn} value (42.1) peaks in the same time window, whereas the highest D_{sn} value related to the minimum dew point temperature (39.6) is achieved on day 29. The classification potential of the dew point temperature to separate the two data sets related to different levels of net blotch risk increases during the growing season until day 40, which can be seen in the ascending trend of all three series in Figure 7.

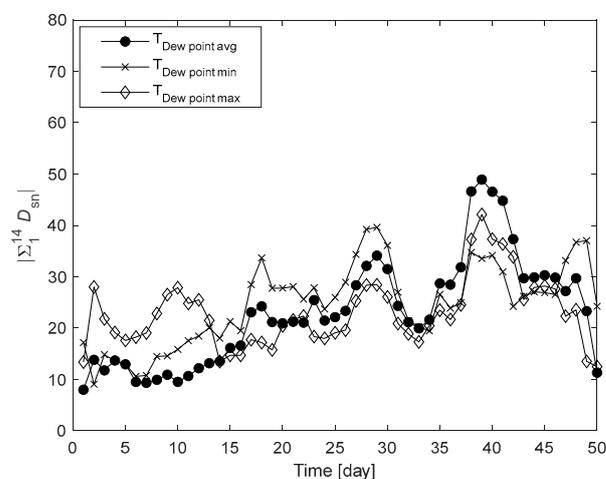


Figure 7. The variation of the classification potential (D_{sn} values) for net blotch risk levels when applying the daily minimum, maximum and average values of the dew point temperature. Day 1 is the beginning of the growing season.

In Figure 8, the D_{sn} values for the daily average, minimum and maximum atmospheric pressure are shown. The highest D_{sn} value, 74.4 is achieved here on day 14 when applying the measured minimum values of the atmospheric pressure. Another peak appears on day 18 and corresponds to a D_{sn} value of 71.4. Furthermore, the highest D_{sn} values of the maximum (49.1; day 13) and average atmospheric pressure (68.4; day 14) peak almost in the same starting time instance. During days 21–28, the D_{sn} values of all three statistical quantities for atmospheric pressure are relatively low. All the D_{sn} values increase slightly after day 30, but are still considerably lower than during days 11–19.

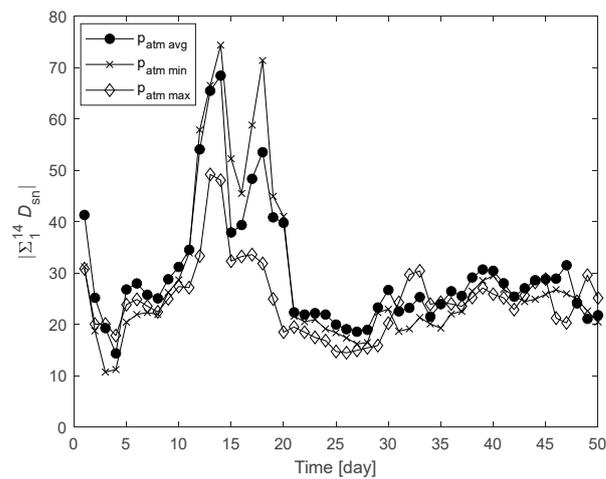


Figure 8. The variation of the classification potential (D_{sn} values) of the daily minimum, maximum and average values of atmospheric pressure. Day 1 is the beginning of the growing season.

In Figure 9, the calculated D_{sn} values for leaf wetness duration (LWD) are presented. The highest D_{sn} value, 34.5 is achieved when the calculation is started on day 22 from the beginning of the growing season. Relatively high D_{sn} values also exist between days 27 and 32. The time window when the maximum D_{sn} values are achieved differs from the peaks presented in Figures 4–7. Here, the maximum D_{sn} value of LWD is at a lower level than the D_{sn} values of the other analysed weather variables.

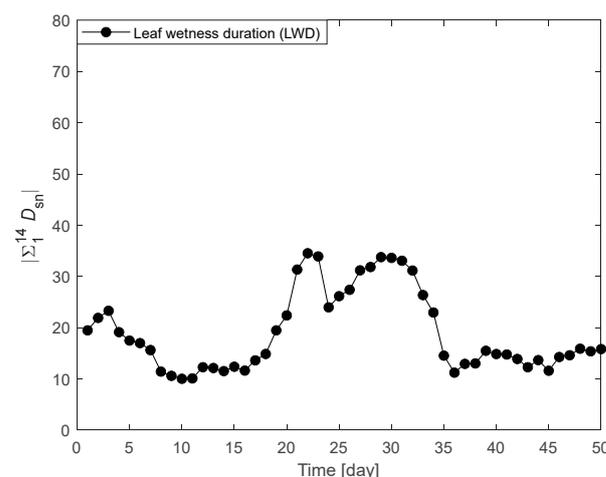


Figure 9. Variation of the classification potential (D_{sn} values) with daily calculated leaf wetness duration. Day 1 on the x -axis is the first day of the growing season.

Figure 10 shows the boxplots of the D_{sn} values for the 715 best-ranked features calculated in data windows of 14 days at each starting time instance (day). On each daily boxplot, the central mark indicates the median of the 715 calculated D_{sn} values of the related features. The bottom and top edges of the boxplot indicate the 25th and 75th percentiles,

respectively. The whiskers extend to the most extreme data points and the individual high D_{sn} values are plotted with 'o' markers.

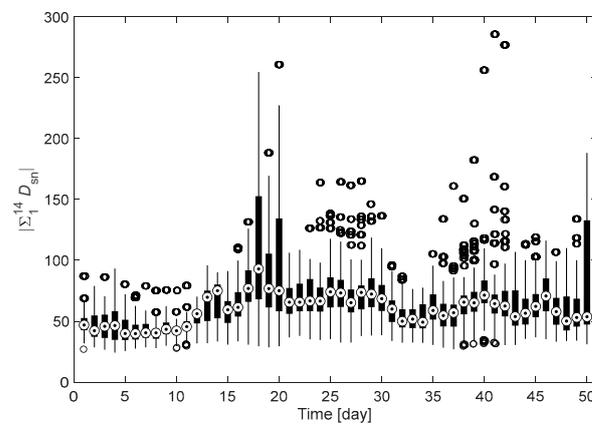


Figure 10. Box plots of the classification potential (D_{sn} values) when using the 115 feature prototypes with 715 different variable combinations at each starting day in time windows of 14 days. The highest individual D_{sn} values are plotted with 'o' markers.

The highest D_{sn} value, namely 285.8, is achieved on day 41 from the beginning of the growing season with the generated feature prototype number 115 (Appendix A), which is structured here as the combination of three weather variables:

$$\frac{\ln(\text{maximum relative humidity})}{\ln(\text{minimum dew point temperature})} \times \ln(\text{minimum outdoor temperature}). \quad (4)$$

These variables included in the feature are also generally known to affect the risk of net blotch infection. The high values are also present on days 20 (D_{sn} value 260.9), 40 (D_{sn} value 256.2) and 42 (D_{sn} value 276.9). The highest statistical median for the D_{sn} value of the plotted features appears on day 18, namely 92.9. It can also be concluded from Figure 10 in comparison to the values of single weather variables (Figures 5–9) that the classification potential generally increases with the applied features. The highest D_{sn} values of analysed variables and of the best generated feature are presented in Table 2.

Table 2. Highest D_{sn} values of analysed variables, related to Figures 5–10.

Variable		Highest D_{sn} Value	Time of the Best D_{sn} Value	Related Figure
Daily outdoor temperature	Avg	52.5	40	5
	Min	29.3	29	
	Max	43.3	40	
Relative humidity	Avg	45.9	25	6
	Min	57	22	
	Max	49.5	49	
Dew point temperature	Avg	48.9	39	7
	Min	39.6	29	
	Max	42.1	39	
Atmospheric pressure	Avg	68.4	14	8
	Min	74.4	14	
	Max	49.1	13	
LWD		34.5	22	9
Feature with the highest D_{sn} value		285.8	41	10

In Figures 11 and 12, the classification potential (sums of 14 days of D_{sn} values) of the best-ranked feature (Equation (2)) is compared at the starting time instances of days 41 and 18 from the beginning of the growing season. The daily D_{sn} values are presented with standard deviations. The markers 'x' and 'o' are the mean values of Category 1 (low net blotch density) and 2 (high net blotch density) data, respectively, and the whiskers describe standard deviation, namely the interval with a confidence level of 68%. In Figure 11, the D_{sn} values with Category 1 data are generally higher than those of Category 2 for every monitored day. Statistically, the categories differ from each other during nine days out of 14 with a confidence level of 68%, as can be seen from Figure 11.

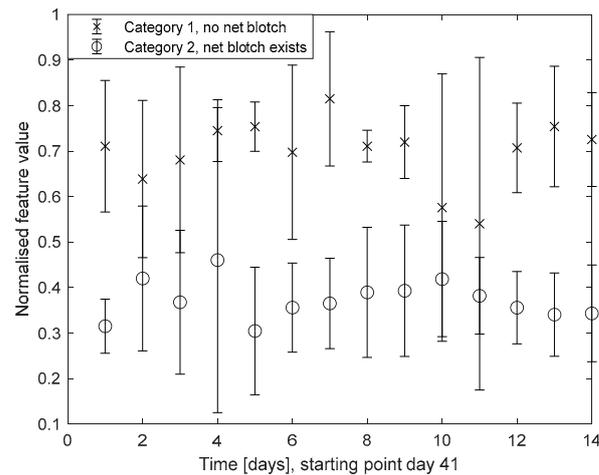


Figure 11. D_{sn} values and their standard deviations with Category 1 and 2 data applying the best feature (Equation (3)) and starting time on day 41 from the beginning of the growing season (14-day data window).

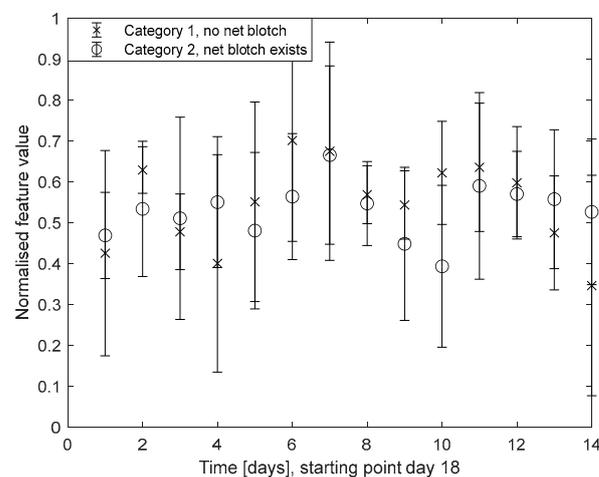


Figure 12. The mean D_{sn} values and their standard deviations with Categories 1 and 2 data sets applying the best feature (Equation (3)) and starting the analysis on day 18 from the beginning of the growing season.

Figure 12 shows that the mean D_{sn} values of Categories 1 and 2 are similar on five days out of 14 and the whiskers overlap in every case, namely with a confidence level of 68%. Thus, statistically the D_{sn} values for the best feature (Equation (2)) are similar in both data sets, leading to poor classification potential.

On the other hand, at least with the generated feature (Equation (4)), it seems that a linear classifier would be sufficient for the task if the analysis starts on day 41 after the beginning of the growing season in Finland. Generally, the results show that the starting

time instance strongly affects the classification potential of net blotch risk levels based on meteorological data.

4. Conclusions

The results show that starting the analysis on day 41 from the beginning of the growing season while applying a 14-day data window would maximise the accuracy of forecasting net blotch risk levels spatially in Finland. The results also indicate that the starting date for forecasting can be identified automatically instead of utilising the sowing date. It can be further concluded that the utilization of features (mathematical transformation of variables) increases the net blotch forecasting potential considerably in comparison to the usage of raw weather variables, including leaf wetness duration. Importantly, it is shown that the selection of an appropriate starting time instance is the crucial factor in developing any forecasting methods for net blotch density, based on information exhibited in meteorological data.

Author Contributions: Conceptualization, O.R.; formal analysis, O.R. and L.J.; methodology, O.R. and M.R.; supervision, K.L.; writing—original draft, O.R., M.J. and L.J.; writing—review and editing, M.J., L.J., M.R. and K.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Agriculture and Forestry of Finland, Document number 632/03.01.02/2017.

Data Availability Statement: See Appendix B.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. The applied feature prototypes.

features(1) = $x - y$;
features(2) = $x - z$;
features(3) = $y - z$;
features(4) = $(x - y) \times y$;
features(5) = $(y - x) \times z$;
features(6) = $(z - x) \times z$;
features(7) = $(y - z) \times z$;
features(8) = $(z - y) \times x$;
features(9) = $(x - z) \times y$;
features(10) = $\ln(x)$;
features(11) = $\ln(y)$;
features(12) = $\ln(z)$;
features(13) = $x \times y$;
features(14) = $x \times z$;
features(15) = $x \times y \times z$;
features(16) = $y \times z$;
features(17) = $\ln(x) - \ln(y)$;
features(18) = $\ln(x) - \ln(z)$;
features(19) = $\ln(y) - \ln(z)$;
features(20) = $\ln(x) - \ln(y) \times \ln(z)$;
features(21) = $\ln(y) - \ln(x) \times \ln(y)$;
features(22) = $\ln(z) - \ln(x) \times \ln(z)$;
features(23) = $\ln(y) - \ln(z) \times \ln(z)$;
features(24) = $\ln(z) - \ln(y) \times \ln(x)$;
features(25) = $\ln(x)/\ln(y)$;
features(26) = $\ln(x) \times \ln(y)$;
features(27) = $\ln(x) \times \ln(z)$;
features(28) = $\ln(x) \times \ln(y) \times \ln(z)$;
features(29) = $\ln(y) \times \ln(z)$;
features(30) = $\text{sqrt}(x)$;

Table A1. *Cont.*

features(31) = \sqrt{y} ;
features(32) = \sqrt{z} ;
features(33) = $\sqrt{x} - \sqrt{y}$;
features(34) = $\sqrt{x} - \sqrt{z}$;
features(35) = $\sqrt{y} - \sqrt{z}$;
features(36) = $\sqrt{\ln(x)}$;
features(37) = $\sqrt{\ln(y)}$;
features(38) = $\sqrt{\ln(z)}$;
features(39) = \sqrt{x}/y ;
features(40) = x/z ;
features(41) = y/z ;
features(42) = $(x \times y)/z$;
features(43) = $(x \times z)/y$;
features(44) = $(y \times z)/x$;
features(45) = \sqrt{x}/\sqrt{y} ;
features(46) = \sqrt{x}/z ;
features(47) = $(y/x)^2$;
features(48) = $(\sqrt{x} \times y)/z$;
features(49) = $(\sqrt{x} \times z)/y$;
features(50) = $(y \times z)/\sqrt{x}$;
features(51) = x^2 ;
features(52) = y^2 ;
features(53) = z^2 ;
features(54) = $x^2 - y^2$;
features(55) = $x^2 - z^2$;
features(56) = x ;
features(57) = y ;
features(58) = z ;
features(59) = $x + y + z$;
features(60) = $x + y - z$;
features(61) = $\ln(x) + \ln(y) + \ln(z)$;
features(62) = $\sqrt{y} + \sqrt{z} + \sqrt{x}$;
features(63) = $(x - y)/x$;
features(64) = $(x/y)^3$;
features(65) = $(y^{(0.7)} - 1)/(0.7)$;
features(66) = $(y - z)/y$; $\%(y - z)/z$, 23.12.2011
features(67) = $(z - y)/x$;
features(68) = $(y^{(-1)} - 1)/(-1)$;
features(69) = $x + y$;
features(70) = $x + z$;
features(71) = $y + z$;
features(72) = $(x + y)/y$;
features(73) = $(y + x)/z$;
features(74) = $(y^{(0.5)} - 1)/(0.5)$;
features(75) = $(z^{(2.5)} - 1)/(2.5)$;
features(76) = $(z + y)/x$;
features(77) = $(y^{(1.5)} - 1)/(1.5)$;
features(78) = $(x + z)/x$;
features(79) = $(y^{(-2)} - 1)/(-2)$;
features(80) = $(x + z)/y$;
features(81) = $\ln(x) + \ln(y)$;
features(82) = $\ln(x) + \ln(z)$;
features(83) = $\ln(y) + \ln(z)$;
features(84) = $(\ln(x) + \ln(y)) \times \ln(z)$;
features(85) = $(\ln(y) + \ln(x)) \times \ln(y)$;
features(86) = $(\ln(z) + \ln(x)) \times \ln(z)$;
features(87) = $(\ln(y) + \ln(z)) \times \ln(z)$;
features(88) = $(\ln(z) + \ln(y)) \times \ln(x)$;
features(89) = $(\ln(x) + \ln(z)) \times \ln(y)$;
features(90) = $\sqrt{x} + \sqrt{y}$;

Table A1. Cont.

```

features(91) = sqrt(x) + sqrt(z);
features(92) = sqrt(y) + sqrt(z);
features(93) = (x + y) × y;
features(94) = (y + x) × z;
features(95) = (z + x) × z;
features(96) = (y + z) × z;
features(97) = (z + y) × x;
features(98) = (x + z) × y;
features(99) = (x + z) × x;
features(100) = (x − y) × x;
features(101) = x + (y × y);
features(102) = y + (x × z);
features(103) = z + (x × z);
features(104) = y + (z × z);
features(105) = z + (y × x);
features(106) = x + (z × y);
features(107) = x + (z × x);
features(108) = x − (y × x);
features(109) = y2 − z2;
features(110) = x2 × y2;
features(111) = (x − y) × z;
features(112) = (x + y) × z;
features(113) = (x/y) × z;
features(114) = (x/y) + z;
features(115) = ln(x)/ln(y) × ln(z);

```

Where x, y and z are the three selected weather quantities for each tested variable combination.

Appendix B

Table A2. Data sources.

The weather data used has been downloaded from the fmi open database:
<https://www.ilmatieteenlaitos.fi/havaintojen-lataus#!/> (accessed on 2 October 2022)

Mynämäki: until 2011, the FMI weather station “Turku airport” and 2012–2017 the FMI weather station “Kaarina, Yltöinen”.
 Jokioinen: the FMI weather station “Jokioinen”.
 Seinäjoki: the FMI weather station “Seinäjoki, Pelmaa”.
 Siikajoki: the FMI weather station “Siikajoki, Revonlahti”.

See the selected data sets and years from Table 1.

References

1. Ruusunen, O.; Jalli, M.; Jauhiainen, L.; Ruusunen, M.; Leiviskä, K. Data Analysis in Moving Windows for Optimizing Barley Net Blotch Prediction. *J. Adv. Agric. Technol.* **2020**, *7*, 154–196. [CrossRef]
2. FAO. FAOSTAT. 2020. Available online: <http://www.fao.org/faostat/en/> (accessed on 21 January 2022).
3. Jalli, M.; Laitinen, P.; Latvala, S. The emergence of cereal fungal diseases and the incidence of leaf spot diseases in Finland. *Agric. Food Sci.* **2011**, *20*, 62–73. [CrossRef]
4. Jalli, M.; Kaseva, J.; Andersson, B.; Ficke, A.; Nistrup-Jørgensen, L.; Ronis, A.; Kaukoranta, T.; Ørum, J.-E.; Djurle, A. Yield increases due to fungicide control of leaf blotch diseases in wheat and barley as a basis for IPM decision-making in the Nordic-Baltic region. *Eur. J. Plant Pathol.* **2020**, *158*, 315–333. [CrossRef]
5. Teferi, T.A.; Wubshet, M.L.; Aregawi, T.B. Occurrence and intensity of net and spot blotch of barley in South Tigray, Ethiopia. *Glob. Sci. Res. J.* **2015**, *3*, 113–123.
6. Agriculture Victoria Net blotches of barley. 2020. Available online: <https://agriculture.vic.gov.au/biosecurity/plant-diseases/grain-pulses-and-cereal-diseases/net-blotches-of-barley> (accessed on 1 September 2022).
7. El Yousfi, B.; Ezzahiri, B. Net Blotch on semi-arid regions of Morocco II—Yield and yield-loss modelling. *Field Crops Res.* **2002**, *73*, 81–93. [CrossRef]
8. Jayasena, K.W.; Van Burgel, C.A.; Tanaka, K.; Majewski, J.; Loughman, R. Yield reduction in barley in relation to spot-type net blotch. *Australas. Plant Pathol.* **2007**, *36*, 429–433. [CrossRef]

9. Turkington, T.K.; Tekauz, A.; Xi, K.; Kutcher, H.R. Foliar diseases of barley: Don't rely on a single strategy from the disease management toolbox. *Prairies Soils Crops J.* **2011**, *4*, 142–150.
10. Aktar, W.; Sengupta, D.; Chowdhury, A. Impact of pesticides use in agriculture: Their benefits and hazards. *Interdiscip. Toxicol.* **2009**, *2*, 1–12. [[CrossRef](#)]
11. European Union. Directive 2009/128/EC of the European Parliament and the Council of 21 October 2009: Establishing a Framework for Community Action to Achieve the Sustainable use of Pesticides. *Off. J. Eur. Union* **2009**, *309*, 71–86.
12. European Commission. Green Deal: Pioneering Proposals to Restore Europe's Nature by 2050 and Halve Pesticide Use by 2030. 2022. Available online: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_3746 (accessed on 1 September 2022).
13. Charaya, M.U.; Upadhyay, A.; Bhati, H.P.; Kumar, A. Plant disease forecasting: Past practices to emerging technologies. In *Plant Disease: Management Strategies*; Nehra, S., Ed.; Agrobios Research: Rajasthan, India, 2021; pp. 1–30.
14. Fenu, G.; Mallocci, F.M. Forecasting Plant and Crop Disease: An Explorative Study on Current Algorithms. *Big Data Cogn. Comput.* **2021**, *5*, 2. [[CrossRef](#)]
15. Gent, D.H.; Mahaffee, W.F.; McRoberts, N.; Pfender, W.F. The Use and Role of Predictive Systems in Disease Management. *Annu. Rev. Phytopathol.* **2013**, *51*, 267–289. [[CrossRef](#)] [[PubMed](#)]
16. Jørgensen, L.N.; Matzen, N.; Ficke, A.; Andersson, B.; Jalli, M.; Ronis, A.; Nielsen, G.C.; Erlund, P.; Djurle, A. Using risk models for control of leaf blotch diseases in barley minimises fungicide use—Experiences from the Nordic and Baltic countries. *Acta Agric. Scand. Sect. B Soil Plant Sci.* **2021**, *71*, 247–260. [[CrossRef](#)]
17. Secher, B.J.M.; Jørgensen, L.N.; Murali, N.S.; Boll, P.S. Field validation of a decision support system for the control of pests and diseases in cereals in denmark. *Pestic. Sci.* **1995**, *45*, 195–199. [[CrossRef](#)]
18. Henriksen, K.E.; Jørgensen, L.N.; Nielsen, G.C. PC-plant protection—A tool to reduce fungicide input in winter wheat, winter barley and spring barley in Denmark. In Proceedings of the Brighton Crop Protection Conference—Pest and Diseases, Brighton, UK, 13–16 November 2000; pp. 835–840.
19. Bligaard, J.; Jørgensen, L.N.; Axelsen, J.; Hansen, J.G.; Ørum, J.E.; Baby, S.; Nielsen, G.C. *Udvikling af Nye Risikomodeller for Septoria (Zymoseptoria tritici) i Vinterhvede*; Miljø-og Fødevarerministeriet, Miljøstyrelsen. Bekæmpelsesmiddelforskning: Odense, Denmark, 2017; p. 168. ISBN 978-87-93529-68-7.
20. WisuEnnuste. 2022. Available online: <https://www.minunmaatilani.fi/ohjelmistot-ja-palvelut/viljelysuunnitteluohjelmat/wisuennuste-kasvinsuojelun-tasmalliseen-ajoittamiseen> (accessed on 3 February 2022).
21. El Jarroudi, M.; Kouadio, A.L.; El Jarroudi, M.; Junk, J.; Bock, C.; Diouf, A.A.; Delfosse, P. Improving fungal disease forecasts in winter wheat: A critical role of intra-day variations of meteorological conditions in the development of Septoria leaf blotch. *Field Crops Res.* **2017**, *213*, 12–20. [[CrossRef](#)]
22. Fernando, W.G.D.; Oghenekaro, A.O.; Tucker, J.R.; Badea, A. Building on a foundation: Advances in epidemiology, resistance breeding, and forecasting research for reducing the impact of fusarium head blight in wheat and barley. *Can. J. Plant Pathol.* **2021**, *43*, 495–526. [[CrossRef](#)]
23. Landschoot, S.; Waegeman, W.; Audenaert, K.; Van Damme, P.; Vandepitte, J.; De Baets, B.; Haesaert, G. A field-specific web tool for the prediction of Fusarium head blight and deoxynivalenol content in Belgium. *Comput. Electron. Agric.* **2013**, *93*, 140–148. [[CrossRef](#)]
24. Musa, T.; Hecker, A.; Vogelgsang, S.; Forrer, H.R. Forecasting of Fusarium head blight and deoxynivalenol content in winter wheat with FusaProg. *EPPPO Bull.* **2007**, *37*, 283–289. [[CrossRef](#)]
25. Shah, D.A.; Paul, P.A.; De Wolf, E.D.; Madden, L.V. Predicting plant disease epidemics from functionally-represented weather series. *Phil. Trans. R. Soc. B* **2019**, *374*, 20180273. [[CrossRef](#)]
26. Shah, D.A.; De Wolf, E.D.; Paul, P.A.; Madden, L.V. Functional data analysis of weather variables linked to Fusarium head blight epidemics in the United States. *Phytopathology* **2019**, *109*, 96–110. [[CrossRef](#)]
27. Shah, D.A.; De Wolf, E.D.; Paul, P.A.; Madden, L.V. Predicting Fusarium head blight epidemics with boosted regression tree. *Phytopathology* **2014**, *104*, 702–714. [[CrossRef](#)]
28. Shah, D.A.; Molineros, J.E.; Paul, P.A.; Willyerd, K.T.; Madden, L.V.; De Wolf, E.D. Predicting Fusarium head blight epidemics with weather-driven pre- and post-anthesis logistic regression model. *Phytopathology* **2013**, *103*, 906–919. [[CrossRef](#)]
29. Ruusunen, O.; Jalli, M.; Jauhiainen, L.; Ruusunen, M.; Leiviskä, K. Advanced Data Analysis as a Tool for Net Blotch Density Estimation in Spring Barley. *Agriculture* **2020**, *10*, 179. [[CrossRef](#)]
30. Saari, E.E.; Prescott, M. A scale for appraising the foliar intensity of wheat diseases. *Plant Dis. Rep.* **1975**, *59*, 377–379.
31. Blum, A.L.; Langley, P. Selection of relevant features and examples in machine learning. *Artif. Intell.* **1997**, *97*, 245–271. [[CrossRef](#)]
32. Dash, M.; Liu, H. Feature selection for classification. *Intell. Data Anal.* **1997**, *1*, 131–156. [[CrossRef](#)]
33. García-Torres, M.; Gómez-Vela, F.; Melián-Batista, B.; Moreno-Vega, J.M. High-dimensional feature selection via feature grouping: A Variable Neighborhood Search approach. *Inf. Sci.* **2016**, *326*, 102–118. [[CrossRef](#)]
34. Pérez-Rodríguez, J.; Arroyo-Peña, A.G.; García-Pedrajas, N. Simultaneous instance and feature selection and weighting using evolutionary computation: Proposal and study. *Appl. Soft Comput.* **2015**, *37*, 416–443. [[CrossRef](#)]
35. Uncu, Ö.; Türkşen, I.B. A novel feature selection approach: Combining feature wrappers and filters. *Inf. Sci.* **2007**, *177*, 449–466. [[CrossRef](#)]
36. Ruusunen, M. Signal Correlations in Biomass Combustion—An Information Theoretic Analysis. *Acta Univ. Ouluensis Ser. C* **2013**, *459*, 1–120.

37. Kruit, R.J.W.; van Pul, W.A.J.; Jacobs, A.F.G.; Heusinkveld, B.G. Comparison between four methods to estimate leaf wetness caused by dew on grassland. In Proceedings of the 26th Conference on Agricultural and Forest Meteorology, Session 10.1, Vancouver, BC, Canada, 23–26 August 2004.
38. Pomeroy, S.; Tamayo, P.; Gaasenbeek, M.; Sturla, L.M.; Angelo, M.; McLaughlin, M.E.; Kim, J.Y.H.; Goumnerovak, L.C.; Blackk, P.M.; Lau, C.; et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* **2002**, *415*, 436–442. [[CrossRef](#)] [[PubMed](#)]