



Article

---

# Linear Discriminant Analysis for Predicting Net Blotch Severity in Spring Barley with Meteorological Data in Finland



---

Outi Ruusunen, Marja Jalli, Lauri Jauhiainen, Mika Ruusunen and Kauko Leiviskä



## Article

# Linear Discriminant Analysis for Predicting Net Blotch Severity in Spring Barley with Meteorological Data in Finland

Outi Ruusunen <sup>1,\*</sup>, Marja Jalli <sup>2</sup>, Lauri Jauhiainen <sup>2</sup>, Mika Ruusunen <sup>1</sup> and Kauko Leiviskä <sup>1</sup>

<sup>1</sup> Control Engineering, Environmental and Chemical Engineering Research Unit, University of Oulu, 90014 Oulu, Finland; mika.ruusunen@oulu.fi (M.R.); kauko.leiviska@oulu.fi (K.L.)

<sup>2</sup> Natural Resources Institute Finland, 31600 Jokioinen, Finland; marja.jalli@luke.fi (M.J.); lauri.jauhiainen@luke.fi (L.J.)

\* Correspondence: outi.ruusunen@oulu.fi

**Abstract:** Predictive information on plant diseases could help to reduce and optimize the usage of pesticides in agriculture. This research presents classification procedures with linear discriminant analysis to predict three possible severity levels of net blotch in spring barley in Finland. The weather data utilized for classification included mathematical transformations, namely features of outdoor temperature and air humidity with calculated dew point temperature values. Historical field observations of net blotch density were utilized as a target class for the identification of classifiers. The performance of classifiers was analyzed in sliding data windows of two weeks with selected, cumulative, summed feature values. According to classification results from 36 yearly data sets, the prediction of net blotch occurrence in spring barley in Finland can be considered as a linearly separable classification task. Furthermore, this can be achieved with linear discriminant analysis by combining the output probabilities of separate binary classifiers identified for each severity level of net blotch disease. In this case, perfect classification with a resolution of three different net blotch severity levels was achieved during the first 50 days from the beginning of the growing season. This strongly suggests that real-time classification based on a few weather variables measured on a daily basis can be applied to estimate the severity of net blotch in advance. This allows application of the principles of integrated pest management (IPM) and usage of pesticides only when there is a proven need.

**Keywords:** barley; net blotch; classification model; decision support system



**Citation:** Ruusunen, O.; Jalli, M.; Jauhiainen, L.; Ruusunen, M.; Leiviskä, K. Linear Discriminant Analysis for Predicting Net Blotch Severity in Spring Barley with Meteorological Data in Finland. *Agriculture* **2024**, *14*, 1779. <https://doi.org/10.3390/agriculture14101779>

Received: 16 August 2024

Revised: 24 September 2024

Accepted: 8 October 2024

Published: 9 October 2024



**Copyright:** © 2024 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 growing global demand for nutrition and the need for a sustainable and robust food chain set increasing requirements on modern agriculture. The European agriculture and food system must follow the Green Deal principles that oblige EU members to find sustainable solutions for farming. One of the targets is for farmers to optimize chemical crop protection and minimize the use of pesticides, with the aim of reducing the use and risk of chemical pesticides by 50% by 2030 [1]. The early detection of plant diseases has an important role in crop protection. This research presents a data-based classification approach for predicting the severity of net blotch (*Pyrenophora teres*), which is one of the main foliar diseases [2] in spring barley (*Hordeum vulgare*), limiting its yield.

Progress in computing capacity has enabled the use of complex plant disease detection systems. Some automated approaches for solving the plant disease detection problem are presented in [3–6]. In the presented cases, plant diseases are identified based on images of infected and healthy plants with the utilization of neural networks or deep learning. Plant disease detection using hyperspectral data and neural networks has been reviewed by Golhani et al. [7], while Whetton et al. [8,9] have utilized hyperspectral data to measure yellow rust and fusarium head blight in wheat and barley. The research and data collection

were performed first in a laboratory [8], after which on-line field measurements were implemented [9]. In an article by Kamilaris and Prenafeta-Boldú [10], a survey of deep learning in agriculture was performed and documented with several examples. The authors discuss that most of the deep learning studies they reviewed were related to computer vision and image analysis [10]. Deep learning has also been utilized in plant disease and species detection systems as presented by Keceli et al. [11]. The authors apply a multi-input network that uses raw images and transferred deep features extracted from a pre-trained deep model to predict plant type and disease, employing a so-called multi-task prediction system [11]. Plant disease identification and classification by combining the red deer optimization algorithm and a deep learning convolutional neural network (ResNet-50) have been presented by Reddy et al. [12]. The authors used a CNN classifier model to build a classification model to predict plant diseases from a plant image [12]. Studies on plant leaf disease detection based on image analysis are also presented in [13–16]. However, these sophisticated plant disease detection systems may require extensive new measurements and real-time data transfer facilities. In contrast, this research on plant disease prediction relies on public weather information and available spatial reference data.

Earlier studies have shown that weather conditions and especially leaf wetness duration are often considered when estimating the risk of plant diseases. In [17,18], a weather data-based prediction system for net blotch occurrence in spring barley together with the most commonly used variables for prediction is discussed in detail. The classification of two different net blotch severity levels (values of 0.5% and 0.6–5%) has been studied by Ruusunen et al. [18,19]. In these articles, the selection of variables and classification potential of different weather variables were also studied. Also, in [20], an analysis was performed in sliding data windows to identify the most informative length of time window for the classification task.

In this study, over 26 years of historical net blotch data and weather measurements are combined, pre-processed and applied to selected classifiers as predictors (model inputs). Linear discriminant analysis (LDA) is utilized here as a model structure for the classifiers. LDA has been used earlier in agricultural classification cases, for example, in grain quality traits in rice [21], a potato diversity study [22] and authentication of food products [23]. Here, the input variables for classifiers were selected based on earlier studies by the authors [17–20]. The most suitable feature and the most accurate classification method for three different net blotch severity levels were identified. It was assumed that the most suitable weather variables identified earlier could be adopted because the phenomenon (barley net blotch) was the same; however, the most suitable feature may now differ because the research frame was extended from a two-level classification to a three-level classification. The main hypothesis was that feasible classification results for net blotch severity can be achieved by applying linear discrimination analysis. This hypothesis was tested by building an ensemble of three binary classifiers and one multiclass classifier. As a potential outcome, when the estimation of net blotch severity can be made early in the growing season, it will ensure the need and time for planning plant protection measures.

## 2. Materials and Methods

### 2.1. Data

Two different data sets were employed in this study, i.e., weather measurements and field observations of net blotch density. Weather data are available from the open database of the Finnish Meteorological Institute (FMI) and net blotch observations are from the database of Natural Resources Institute Finland (Luke). Historical net blotch data collected by Luke (1991–2017) was utilized in this research. The data set includes information on the observation year, field location, barley genotype and net blotch severity. Field trials had been conducted by Luke in various locations across Finland's different hardiness zones, with data collection methods following standard Official Variety Trials procedures [24]. The trials included 30 barley cultivars per year, with replicates to ensure robust statistical analysis. Disease pressure was quantified using a statistical model, accounting for environ-

mental effects and genotype variability. For this study, net blotch data was scaled into three categories indicating the amount of net blotch in each data set. In category 1, the maximum net blotch value was 0.5%, and in categories 2 and 3, the net blotch values were 0.6–5% and over 5.1%, respectively. One example of the scale for appraising plant disease severity in cereals is presented in [25]. More about the field experiments and data collection can be found in [18].

The weather measurements were downloaded from the open database and the closest weather stations to the fields were selected from the FMI weather station list [26]. It is worth noting that the net blotch observation data consisted of one value per year while the weather data consisted of daily observations. Information and locations of the weather stations related to the source of meteorological measurements can be found in Appendix A.

The data sets were grouped according to the location of the test fields, observation year and net blotch category (1–3). Data sets with missing information were removed from this study. The beginning of the growing season varies spatially and temporally in Finland, and, therefore, the data sets were pre-processed to set the starting day of each data set as the beginning of the growing season. The beginning of the growing season is determined here as the time when the mean outdoor temperature remains over +5 °C for 10 consecutive days. A detailed description of the weather data and its pre-processing is presented in [18]. All the data analysis and result evaluation were performed with MATLAB® version R2022a.

The locations of the test fields and the years of the selected weather data by net blotch severity categories are presented in Table 1. It also shows the division of overall data into randomly selected (by year) training and test data.

**Table 1.** Description of the randomly selected training and test data sets, including names of the Finnish municipalities where the test fields are located, together with the net blotch severity categories and years corresponding to weather data sets.

Net Blotch Category	Training Data		Test Data	
	Location	Year	Location	Year
1	Hämeenlinna	1991		
	Siikajoki	1992		
	Siikajoki	1993	Seinäjoki	2000
	Jokioinen	2006	Seinäjoki	2007
	Jokioinen	2007		
	Mynämäki	2010		
2	Siikajoki	1991		
	Inkoo	2002		
	Inkoo	2005		
	Siikajoki	2010	Siikajoki	2006
	Mynämäki	2011	Jokioinen	2009
	Seinäjoki	2011	Jokioinen	2010
	Seinäjoki	2013	Jokioinen	2011
	Jokioinen	2013		
Inkoo	2017			
3	Siikajoki	2012	Inkoo	2003
	Mynämäki	2013	Mynämäki	2009
	Mynämäki	2014	Inkoo	2012
	Jokioinen	2014	Loviisa	2014
	Jokioinen	2015	Siikajoki	2014
	Jokioinen	2016	Siikajoki	2015
	Mynämäki	2016	Inkoo	2015
	Seinäjoki	2016		

The downloaded weather data included hourly measured values of the following variables:

- Atmospheric pressure (kPa);
- Relative humidity (RH%);

- Outdoor temperature (°C);
- Dew point temperature (°C).

For this research, the statistical information considered for classification were the daily maximum values of relative humidity, daily minimum of dew point temperatures and daily minimum of outdoor temperatures that have been shown to be significantly related to the occurrence of net blotch severity [19]. The statistical characteristics of these values by each category are presented in Table 2.

**Table 2.** General statistics of the considered weather data quantities calculated for each net blotch category. Data sources for statistical calculations per category, including both training and test data sets, are presented in Table 1.

	Category 1			
	Mean	Standard Deviation	Minimum	Maximum
Daily maximum relative humidity (max RH%)	83.4	16.3	33	100
Daily minimum outdoor temperature (min °C outdoor)	6.2	4.2	−4.6	17.2
Daily minimum dew point temperature (min °C dew point)	3.4	4.7	−11.9	14.3
	Category 2			
	Mean	Standard Deviation	Minimum	Maximum
Daily maximum relative humidity (max RH%)	91.5	9.5	51	100
Daily minimum outdoor temperature (min °C outdoor)	7.5	4.4	−3.7	18.2
Daily minimum dew point temperature (min °C dew point)	4.5	5.0	−13.6	18.2
	Category 3			
	Mean	Standard Deviation	Minimum	Maximum
Daily maximum relative humidity (max RH%)	95.3	6.0	65	100
Daily minimum outdoor temperature (min °C outdoor)	7.5	4.0	−4.5	19
Daily minimum dew point temperature (min °C dew point)	4.9	4.7	−12.9	16.2

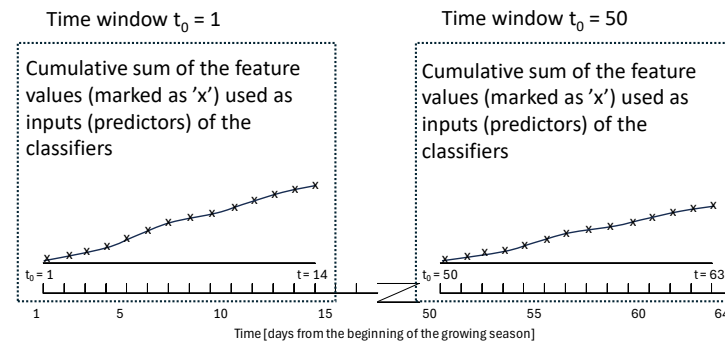
The resulting number of training data points in the following analysis for each considered weather variable was 16,100 and for the tests was 9100 data points, respectively. In total, 75,600 data points were then available for feature generation with the meteorological data and 36 datapoints were available for the target class, namely historical information on net blotch occurrence per year.

### 2.2. Classification Procedure

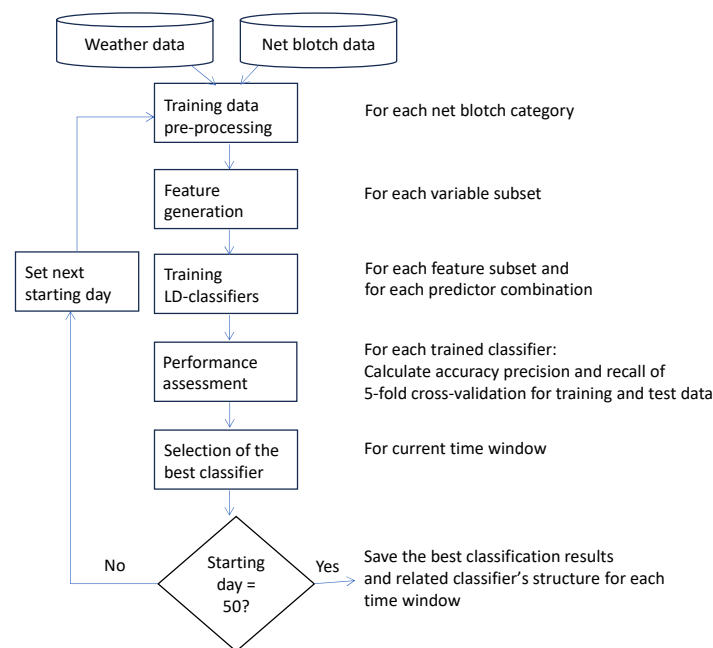
Classification with an ensemble consisting of three linear discriminant (LD) binary classifiers and with one LD multiclass classifier was analyzed in this study. Firstly, the data sets of the net blotch and weather measurements were pre-processed and generated into the features as presented in detail in [18,19]. Then, the cumulative sums of the feature values were computed in a 14-day sliding data window and applied to the classifiers as predictors (inputs). This was repeated in 50 consecutive time windows, with the first data window starting day being the beginning of the growing season as shown in Figure 1. The 14-day time window was chosen based on earlier research [20] as it consistently offers enough information to robustly classify different net blotch severities.

The performance of each constructed classifier (in each 14-day time window) and the most suitable predictors (cumulative summed value of features during days 1–14) was evaluated by accuracy, precision and recall metrics for training and test data. The

classification procedure was applied with the three binary classifiers forming an ensemble classifier and with one multiclass classifier in every 50 data windows. The binary classifiers used were 'Category 1 vs. rest', 'Category 2 vs. rest' and 'Category 3 vs. rest'. The ensemble classifier was generated by aggregating the outputs of those three binary classifiers with arithmetic and geometric averages. The tested classification procedure is presented in Figure 2. This workflow was applied to identify the structure and parameters of the three binary LD-classifiers and one LD-multiclass classifier.



**Figure 1.** Illustration of data analysis in 50 successive sliding time windows through the growing season.



**Figure 2.** Analysis procedure for the tested classifiers.

### 2.2.1. Feature Generation

In [18], it has been presented that the separation ability of the original weather variables is insufficient to predict the appearance of net blotch, and, therefore, feature generation must be performed. In practice, this means that new computational variables were generated from the original data by mathematical operations to reveal the information content of the data. The same feature generation procedure as described in previous studies [18–20] was utilized here for the above-listed variables. The features were generated for the selected variables with different mathematical operations, such as addition, subtraction, multiplication, division, involution, logarithm, square root and combinations of the above. More information about the feature generation methods in general can be found, for example, in [27–31]. The feature generation method used in this study is presented by Ruusunen [32]; the applied feature basis functions for this study are listed in Appendix B.

### 2.2.2. Linear Discriminant Classifiers

Linear discriminant analysis (LDA), which is a supervised learning algorithm and can be used for classification tasks, was utilized for data classification in this research. LDA is a statistical technique based on Fisher's linear discriminant. Details of Fisher's discriminant can be found in [33,34]. LDA can be used to find a linear combination of predictor variables that separates the classes in a data set. With LDA, the data is projected onto a lower-dimensional space that maximizes the between-class distance and minimizes the within-class distance.

Linear discriminant analysis presupposes that each class (Equation (1)) generates data following a Gaussian distribution. Also, it is assumed that each of the classes has the same co-variance matrix. To develop a classifier, the fitting function calculates the parameters of these Gaussian distributions for each class. Once trained, the classifier predicts the classes of new data by identifying the class with the lowest misclassification cost [35]:

$$\hat{y} = \operatorname{argmin}_{y=1\dots K} \sum_{k=1}^K \hat{P}(k|x)C(y|k), \quad (1)$$

where  $\hat{y}$  is the predicted class;  $K$  is the number of classes;  $\hat{P}(k|x)$  is the posterior probability of class  $k$  for observation  $x$ ; and  $\hat{C}(y|k)$  is the cost of classifying an observation as  $y$  when its true class is  $k$ .

For this study, the cost is zero for correct classification and one for incorrect classification.

The posterior probability [35] that data point  $x$  belongs to class  $k$  is calculated as the product of the prior probability of class  $k$  and the value of the multivariate normal density function, namely with the posterior probability presented in Equation (2):

$$\hat{P}(k|x) = \left( \frac{1}{((2\pi)^d |\Sigma_k|)^{1/2}} \right)^{(-\frac{1}{2}(x-\mu_k)\Sigma_k^{-1}(x-\mu_k)^T)} \quad (2)$$

where  $|\Sigma_k|$  is the determinant of  $\Sigma_k$  and  $\Sigma_k^{-1}$  is the inverse matrix.

If  $P(k)$  is the prior probability of class  $k$ , then the posterior probability of an observation  $x$  (of class  $k$ ) is as follows:

$$\hat{P}(k|x) = \frac{P(x|k)P(k)}{P(x)} \quad (3)$$

where  $P(x)$  is the normalization constant, namely, the sum over  $k$  of  $P(x|k)P(k)$ .

### 2.2.3. Classification Model Structures

Two different classification structures, namely an ensemble classifier and multiclass classification, were tested in this study. The ensemble classifier comprises usage of the three binary classifiers for each net blotch category as presented in Figure 3. Particularly, each of the three binary LD classifiers are trained to classify between the net blotch category and the two other categories. These binary classifiers then perform the one-vs.-rest classification task. In the case of the ensemble model, the tested aggregation operations for soft voting with posterior probabilities for each category were the arithmetic mean and geometric mean. Finally, the argmax of the sums of the predicted post probabilities determines the output of the ensemble classifier. The binary classifiers were trained with 23 years of measured data and tested with 13 years of data (see Table 1).

For the multiclass classification, a single classifier was identified for all three classes, namely, the net blotch categories. Here, the function *fitdiscr* in MATLAB<sup>®</sup> was utilized in its default mode. Predictions, namely the posterior probabilities of both classifiers, were achieved using the *predict* function. In the case of the multiclass classifier, the class exhibiting the highest posterior probability was selected as the output of the classifier.

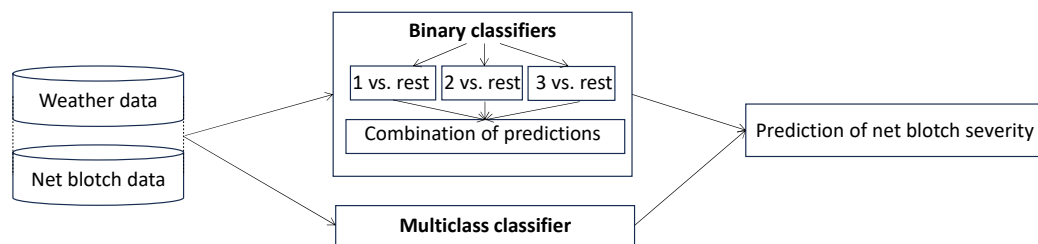


Figure 3. Tested classifier structures.

### 2.2.4. Performance Assessment

Classification performance was assessed with metrics related to precision, recall and accuracy. Precision is the proportion of classifications that are truly correct. It can be computed as the number of positive hits divided by the total number of positively predicted observations [36]:

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

where TP = True Positives and FP = False Positives (observations that are classified as positive but are truly negative).

Recall is the proportion of truly positive identifications divided by the total number of observations belonging to the class. Recall can be computed as [36] follows:

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

where TP = True Positive elements and FN = False Negative (observations that are classified as negative but are truly positive).

Accuracy is a common metric in classification. It presents the proportion of correctly estimated classifications from the total number of classified objects. The equation for accuracy can be presented as [36]

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

where TP = True Positive elements; TN = True Negative elements; FP = False Positive (observations that are classified as positive but are truly negative); and FN = False Negative elements.

## 3. Results

The identified, best-performing features and the optimal number of predictors according to the analysis (Figure 2) are presented in Table 3 when starting day 13 is applied. It is important to notice that the features utilized vary across different data windows in these results. In this study, starting day 13 resulted in the best classification results in general.

Table 3. Final model structures and identified predictors (feature values during selected days in a data window between 1–14) when starting day 13 (from the beginning of the growing season) was applied.

Model Structure	Binary, Category 1	Binary, Category 2	Binary, Category 3	Multiclass
Selected feature	ln(min °C outdoor)	ln(min °C outdoor)	(min °C outdoor–min °C dewpoint)	ln(min °C outdoor)
Selected running numbers of days from the start of a data window	11–14	6–14	1–11	11–14
Number of selected predictors	4	9	11	4



Precision, recall and accuracy were calculated across various time windows for both a multiclass classifier and two distinct ensemble classifiers. The ensemble classifiers differ in their aggregation operation, utilizing both arithmetic and geometric means. The classification results for every time window are presented in Appendices C–E. As an example, the accuracy of the time window with starting day 13 is presented in Figure 4 (multiclass classifier) and Figure 5 (ensemble classifier). The ensemble classifiers with the arithmetic mean and geometric mean as an aggregation method had the same accuracy in the selected time window and, therefore, only one confusion matrix is presented for the ensemble classifiers. In confusion matrices, the classification result (predicted category) is named “Output class” and the real category is named “Target Class”.

		Multiclass classifier			
		Training data		Test data	
Output Class	Target Class	1	2	3	Accuracy
		1	6 26.1%	0 0.0%	0 0.0%
2	0 0.0%	9 39.1%	2 8.7%	81.8% 18.2%	
3	0 0.0%	0 0.0%	6 26.1%	100% 0.0%	
		100% 0.0%	100% 0.0%	75.0% 25.0%	91.3% 8.7%

		Multiclass classifier			
		Training data		Test data	
Output Class	Target Class	1	2	3	Accuracy
		1	1 7.7%	0 0.0%	0 0.0%
2	0 0.0%	3 23.1%	1 7.7%	75.0% 25.0%	
3	1 7.7%	1 7.7%	6 46.2%	75.0% 25.0%	
		50.0% 50.0%	75.0% 25.0%	85.7% 14.3%	76.9% 23.1%

**Figure 4.** Confusion matrix (green for correct, red for incorrect) and accuracy of the multiclass classifier starting on day 13 from the beginning of the growing season.

		Ensemble classifier			
		Training data		Test data	
Output Class	Target Class	1	2	3	Accuracy
		1	6 26.1%	0 0.0%	0 0.0%
2	0 0.0%	9 39.1%	0 0.0%	100% 0.0%	
3	0 0.0%	0 0.0%	8 34.8%	100% 0.0%	
		100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%

		Ensemble classifier			
		Training data		Test data	
Output Class	Target Class	1	2	3	Accuracy
		1	2 15.4%	0 0.0%	0 0.0%
2	0 0.0%	4 30.8%	0 0.0%	100% 0.0%	
3	0 0.0%	0 0.0%	7 53.8%	100% 0.0%	
		100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%

**Figure 5.** Confusion matrix (green for correct, red for incorrect) and overall accuracy of the ensemble classifier with the geometric mean as the aggregation method starting on day 13 from the beginning of the growing season.

The confusion matrix for the multiclass classifier indicates that 21 correct classifications out of a total of 23 instances in the training data were achieved. The matrix reveals two misclassifications where the predicted category was 2, whereas the correct category should have been 3. In the test data, the multiclass classifier achieved 10 correct classifications out of a total of 13. Three misclassifications occurred, with one instance where the predicted class was 2 instead of 3, and two instances where the predicted classes should have been categories 1 and 2, respectively. The overall accuracy of the multiclass classifier was 91.3% on the training data and 76.9% on the test data. The precision and recall for categories 1, 2 and 3 were as follows:

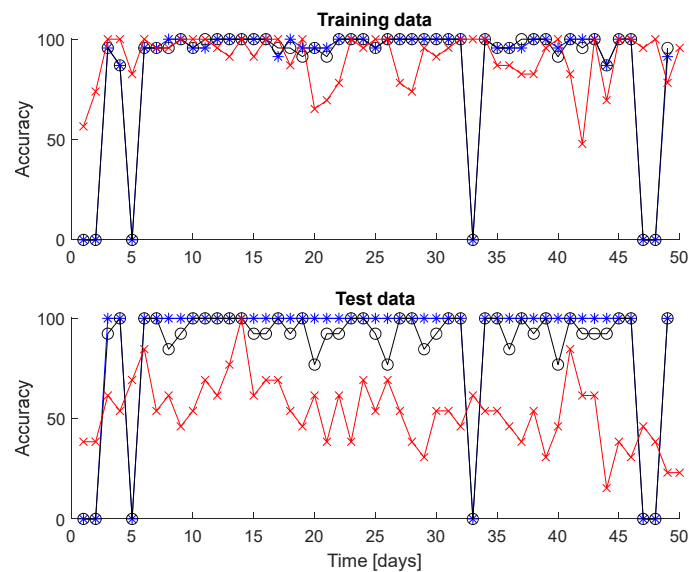
- 100%, 82% and 100% (precision);
- 100%, 100% and 75% (recall).

The precision and recall for the training data and test data were as follows:

- 100%, 75% and 75% (precision);
- 50%, 75% and 85.7% (recall).

The confusion matrix for the ensemble classifier (Figure 4) shows 100% overall accuracy in both training and test data, indicating perfect classification with the available data. During the sliding window starting on day 13, both ensemble classifiers achieved 100% precision and recall in both training and test data.

The overall accuracy of the best-performing classifiers for the whole analyzed time period is presented in Figure 6. The multiclass classifier is denoted by 'x' and a red line, while the ensemble classifiers using the arithmetic mean and geometric mean for the aggregation of binary classifications are represented by 'o' with a black line and '\*' with a blue line, respectively.



**Figure 6.** Overall accuracy of the best-performing ensemble (blue and black o-lines) and multiclass classifiers (red x-line).

#### 4. Discussion

While the ensemble classifier utilizing the geometric mean as the aggregation method consistently exhibits strong performance across almost all considered time windows, the accuracy of the multiclass classifier is comparatively the lowest. Notably, the ensemble classifier with the arithmetic mean achieves 100% overall accuracy for several days. However, it is worth noticing that the prediction results obtained with the geometric mean as the aggregation method in the ensemble classifier are generally better (mostly 100% in overall accuracy with test data) than those with the arithmetic mean. Accuracy of the multiclass classifier reaches its maximum with test data at starting day 14.

According to the results presented in Table 3, the identified final model structures indicate that similar features have been independently found through the analysis procedure. Notably, in every constructed classifier, the optimal feature includes the daily minimum temperature. In three out of four model structures, the mathematical operation used to generate this feature is the natural logarithm (ln). Generally, outdoor temperature is related to dew point temperature and thus to weather conditions affecting the occurrence of net blotch. Additionally, the optimal predictors selected for these classification models are mainly from the end of the two-week period. This result is logical since calculating cumulative sums of time series tends to maximize variation between the resulting data vectors at the end.

A potential uncertainty related to the performance of the classifiers is the distance between weather stations and the fields. Future research aims to quantify this uncertainty by analyzing the effects of location on meteorological measurements concerning barley field measurements. Furthermore, research on the transferability of the presented data-

based methodology to other climate areas is of interest and requires ensuring accurate reference measurements related to the occurrence of net blotch. Additionally, generalizing the presented data-based methodology for predicting other plant diseases that depend on meteorological phenomena is a future interest.

While this research focused on applying variations of linear discriminant classifiers to the problem, many other advanced data-based classification model structures are available, including neural networks (NN), support vector machines (SVM), decision trees, k-nearest neighbors and quadratic discriminant analysis. However, a recent study in [19] has strongly indicated the existence of linear separability for this classification task when utilizing mathematical transformations of meteorological data. Therefore, linear classifiers seem to provide the high accuracy and robustness that are key properties together with applicable model complexity to enable the development of predictive and real-time plant disease warning systems for the future.

Based on the results presented on the prediction of net blotch in spring barley in Finland utilizing meteorological data, the following conclusions can be derived:

1. The results strongly suggest that meteorological data contain sufficient information for the classification of net blotch occurrence in advance during the two weeks from the start of the growing season in Finland.
2. It can be further supposed that the three severity levels of net blotch studied here are linearly separable when the classification is based on mathematical features derived from meteorological data applying linear discriminant analysis.
3. Based on the analyzed data, the ensemble of discriminant classifiers generally outperformed the single multiclass classifier in terms of overall accuracy.
4. In ensemble classification and with the considered predictors, the suggested aggregation method for this purpose is the geometric mean.
5. By utilizing linear discriminant analysis, an early warning system for barley net blotch severity can be implemented using public weather data as the input.

Future research will focus on further validation of the resulting predictive modelling approach with new measurements. This will also involve research on the easy applicability and automatic calibration of the proposed methodology.

**Author Contributions:** Conceptualization, O.R.; formal analysis, O.R.; methodology, O.R. and M.R.; supervision, K.L.; writing—original draft, O.R.; 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.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A

### Data Sources

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

Weather stations nearby test fields are located as follows:

Hämeenlinna: “Hämeenlinna Lammi Pappila” FMI weather station

Location: Latitude 61.05408, Longitude. 25.03844

Inkoo: “Inkoo Bågaskär” FMI weather station

Location: Latitude 59.93114, Longitude 24.01408

Loviisa: “Porvoo Emäsalo” FMI weather station

Location: Latitude 60.20382, Longitude 25.62546

Mynämäki: “Turku airport” FMI weather station until 2011 and “Kaarina, Yltöinen” FMI weather station 2012–2017.

Location: Latitude 60.51565, Longitude 22.27916

Jokioinen: “Jokioinen” FMI weather station.

Location: Latitude 60.81397, Longitude 23.49825

Seinäjäki: “Seinäjäki, Pelmaa” FMI weather station.

Location: Latitude 62.93808, Longitude 22.48878

Siikajoki: “Siikajoki, Revonlahti” FMI weather station.

Location: Latitude 64.68421, Longitude 25.08919

Please see the selected data sets and years in Table 1

## Appendix B

### Feature Basis Functions

- 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) =  $\sqrt{x}$ ;
- 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) =  $(y - z)/y$ ;  
 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) =  $\sqrt{x} + \sqrt{y}$ ;  
 features (91) =  $\sqrt{x} + \sqrt{z}$ ;  
 features (92) =  $\sqrt{y} + \sqrt{z}$ ;  
 features (93) =  $(x + y) \times y$ ;  
 features (94) =  $(y + x) \times z$ ;  
 features (95) =  $(z + x) \times z$

features (96) =  $(y + z) \times x$ ;  
 features (97) =  $(z + y) \times x$ ;  
 features (98) =  $(x + z) \times y$ ;  
 features (99) =  $(x + z) \times x$ ;  
 features (100) =  $(x - y) \times x$ ;  
 features (101) =  $x + (y \times y)$ ;  
 features (102) =  $y + (x \times z)$ ;  
 features (103) =  $z + (x \times z)$ ;  
 features (104) =  $y + (z \times z)$ ;  
 features (105) =  $z + (y \times x)$ ;  
 features (106) =  $x + (z \times y)$ ;  
 features (107) =  $x + (z \times x)$ ;  
 features (108) =  $x - (y \times x)$ ;  
 features (109) =  $y^2 - z^2$ ;  
 features (110) =  $x^2 \times y^2$ ;  
 features (111) =  $(x - y) \times z$ ;  
 features (112) =  $(x + y) \times z$ ;  
 features (113) =  $(x./y) \times z$ ;  
 features (114) =  $(x./y) + z$ ;  
 features (115) =  $\ln(x)/\ln(y) \times \ln(z)$ ;  
 where  $x$ ,  $y$  and  $z$  are the three selected weather quantities for each tested variable combination.

### Appendix C

Table A1. Results of the Multiclass classifier.

Training Data						Test Data					
Precision			Recall			Precision			Recall		
Category 1	Category 2	Category 3	Category 1	Category 2	Category 3	Category 1	Category 2	Category 3	Category 1	Category 2	Category 3
0.80	0.60	0.38	0.67	0.67	0.38	1.00	0.29	0.40	0.50	0.50	0.29
0.83	0.78	0.63	0.83	0.78	0.63	1.00	0.00	0.44	0.50	0.00	0.57
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.33	0.63	1.00	0.25	0.71
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.36	1.00	0.50	1.00	0.14
0.60	0.90	0.75	0.50	1.00	0.75	1.00	0.50	0.71	1.00	0.50	0.71
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.75	0.86	1.00	0.75	0.86
1.00	0.90	1.00	1.00	1.00	0.88	1.00	0.33	0.56	0.50	0.25	0.71
1.00	0.90	1.00	1.00	1.00	0.88	1.00	0.25	0.63	0.50	0.25	0.71
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.38	0.50	0.50	0.75	0.29
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.29	0.60	0.50	0.50	0.43
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.64	0.50	0.25	1.00
1.00	0.90	1.00	1.00	1.00	0.88	1.00	0.40	1.00	0.50	1.00	0.29
1.00	0.82	1.00	1.00	1.00	0.75	1.00	0.75	0.75	0.50	0.75	0.86
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.80	1.00	0.50	1.00	1.00
1.00	1.00	0.80	1.00	0.78	1.00	1.00	0.33	0.67	0.50	0.50	0.57
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.33	0.67	0.50	0.25	0.86
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.60	0.71	0.50	0.75	0.71
1.00	0.80	0.86	1.00	0.89	0.75	1.00	0.40	1.00	1.00	1.00	0.14
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.50	1.00	0.00	0.57
0.83	0.71	0.50	0.83	0.56	0.63	1.00	0.38	0.75	0.50	0.75	0.43
1.00	0.71	0.55	0.83	0.56	0.75	1.00	0.22	0.33	0.50	0.50	0.14
1.00	0.78	0.67	0.83	0.78	0.75	1.00	0.40	0.67	1.00	0.50	0.57



Table A2. Cont.

Training Data						Test Data					
Precision			Recall			Precision			Recall		
Category 1	Category 2	Category 3	Category 1	Category 2	Category 3	Category 1	Category 2	Category 3	Category 1	Category 2	Category 3
1.00	0.90	0.89	0.67	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	0.90	1.00	1.00	1.00	0.88	1.00	0.80	0.86	0.50	1.00	0.86
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.80	1.00	1.00	1.00	0.86
1.00	0.90	1.00	1.00	1.00	0.88	1.00	0.80	1.00	0.50	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.80	1.00	1.00	1.00	0.86
1.00	0.90	1.00	0.83	1.00	1.00	1.00	0.75	0.88	0.50	0.75	1.00
1.00	1.00	0.89	1.00	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	0.89	1.00	0.89	1.00	1.00	0.80	1.00	1.00	1.00	0.86
1.00	0.89	0.88	1.00	0.89	0.88	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	0.89	0.83	1.00	1.00	1.00	0.57	1.00	1.00	1.00	0.57
0.80	0.80	0.88	0.67	0.89	0.88	1.00	0.80	1.00	1.00	1.00	0.86
1.00	0.90	1.00	0.83	1.00	1.00	1.00	1.00	0.88	1.00	0.75	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	0.82	1.00	0.83	1.00	0.88	1.00	0.75	0.88	0.50	0.75	1.00
1.00	1.00	1.00	1.00	1.00	1.00	0.50	0.40	0.83	0.50	0.50	0.71
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	0.67	1.00	1.00	1.00	0.75	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.67	1.00	1.00	1.00	0.71
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.75	0.88	0.50	0.75	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	0.67	1.00	1.00	1.00	0.75	1.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	0.89	0.83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	0.89	1.00	0.89	1.00	1.00	0.75	0.86	1.00	0.75	0.86
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.80	1.00	1.00	1.00	0.86
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	0.80	0.83	0.89	1.00	0.67	0.75	0.83	1.00	0.75	0.71
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	0.90	0.89	0.67	1.00	1.00	0.00	0.50	1.00	0.00	0.50	0.86
1.00	1.00	1.00	1.00	1.00	1.00	0.50	0.80	1.00	0.50	1.00	0.86
1.00	1.00	0.73	0.67	0.89	1.00	1.00	1.00	0.88	1.00	0.75	1.00





Table A3. Cont.

Training Data						Test Data					
Precision			Recall			Precision			Recall		
Category 1	Category 2	Category 3	Category 1	Category 2	Category 3	Category 1	Category 2	Category 3	Category 1	Category 2	Category 3
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	0.89	0.83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	0.89	1.00	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	0.89	1.00	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	0.89	0.83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	0.90	1.00	0.83	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	0.73	0.67	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.00	1.00	0.80	1.00	0.78	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

## References

1. European Commission. The Farm to Fork Strategy. Available online: [https://food.ec.europa.eu/horizontal-topics/farm-fork-strategy\\_en](https://food.ec.europa.eu/horizontal-topics/farm-fork-strategy_en) (accessed on 1 July 2024).
2. 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. [\[CrossRef\]](#)
3. Argüeso, D.; Picon, A.; Irusta, U.; Medela, A.; San-Emeterio, M.G.; Bereciartua, A.; Alvarez-Gila, A. Few-Shot Learning approach for plant disease classification using images taken in the field. *Comput. Electron. Agric.* **2020**, *175*, 105542. [\[CrossRef\]](#)
4. Chen, J.; Chen, J.; Zhang, D.; Sun, Y.; Nanekaran, Y. Using deep transfer learning for image-based plant disease identification. *Comput. Electron. Agric.* **2020**, *173*, 105393. [\[CrossRef\]](#)
5. Ferentinos, K.P. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* **2018**, *145*, 311–318. [\[CrossRef\]](#)
6. Tiwari, V.; Joshi, R.C.; Dutta, M.K. Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images. *Ecol. Inform.* **2021**, *63*, 101289. [\[CrossRef\]](#)
7. Golhani, K.; Balasundram, S.K.; Vadamalai, G.; Pradhan, B. A review of neural networks in plant disease detection using hyperspectral data. *Inf. Process. Agric.* **2018**, *5*, 354–371. [\[CrossRef\]](#)

8. Whetton, R.L.; Hassall, K.L.; Waine, T.W.; Mouazen, A.M. Hyperspectral measurements of yellow rust and fusarium head blight in cereal crops: Part 1: Laboratory study. *Biosyst. Eng.* **2018**, *166*, 101–115. [[CrossRef](#)]
9. Whetton, R.L.; Waine, T.W.; Mouazen, A.M. Hyperspectral measurements of yellow rust and fusarium head blight in cereal crops: Part 2: On-line field measurement. *Biosyst. Eng.* **2018**, *167*, 144–158. [[CrossRef](#)]
10. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [[CrossRef](#)]
11. Keceli, A.S.; Kaya, A.; Catal, C.; Tekinerdogan, B. Deep learning-based multi-task prediction system for plant disease and species detection. *Ecol. Inform.* **2022**, *69*, 101679. [[CrossRef](#)]
12. Reddy, S.R.; Varma, G.S.; Davuluri, R.L. Resnet-based modified red deer optimization with DLCNN classifier for plant disease identification and classification. *Comput. Electr. Eng.* **2023**, *105*, 108492. [[CrossRef](#)]
13. Sharma, V.; Tripathi, A.K.; Mittal, H. DLNC-Net: Deeper lightweight multi-class classification model for plant leaf disease detection. *Ecol. Inform.* **2023**, *75*, 102025. [[CrossRef](#)]
14. Jackulin, C.; Murugavalli, S. A comprehensive review on detection of plant disease using machine learning and deep learning approaches. *Meas. Sens.* **2022**, *24*, 100441. [[CrossRef](#)]
15. Sutaji, D.; Yıldız, O. LEMOXINET: Lite ensemble MobileNetV2 and Xception models to predict plant disease. *Ecol. Inform.* **2022**, *70*, 101698. [[CrossRef](#)]
16. Iqbal, Z.; Khan, M.A.; Sharif, M.; Shah, J.H.; ur Rehman, M.H.; Javed, K. An automated detection and classification of citrus plant diseases using image processing techniques: A review. *Comput. Electron. Agric.* **2018**, *153*, 12–32. [[CrossRef](#)]
17. Mäyrä, O.; Ruusunen, M.; Jalli, M.; Jauhiainen, L.; Leiviskä, K. Plant Disease Outbreak—Prediction by Advanced Data Analysis. *SNE Simul. Notes Eur.* **2018**, *28*, 113–115. [[CrossRef](#)]
18. 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](#)]
19. 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. [[CrossRef](#)]
20. 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*, 38–42. [[CrossRef](#)]
21. Deepika, C.; Gnanamalar, R.P.; Thangaraj, K.; Revathy, N.; Karthikeyan, A. Linear Discriminant Analysis of Grain Quality Traits in Rice (*Oryza sativa* L.) Using the Digital Imaging Technique. *J. Cereal Sci.* **2023**, *109*, 103609. [[CrossRef](#)]
22. Sampaio, S.L.; Barreira, J.C.; Fernandes, Â.; Petropoulos, S.A.; Alexopoulos, A.; Santos-Buelga, C.; Ferreira, I.C.; Barros, L. Potato Biodiversity: A Linear Discriminant Analysis on the Nutritional and Physicochemical Composition of Fifty Genotypes. *Food Chem.* **2021**, *345*, 128853. [[CrossRef](#)]
23. Esteki, M.; Shahsavari, Z.; Simal-Gandara, J. Use of Spectroscopic Methods in Combination with Linear Discriminant Analysis for Authentication of Food Products. *Food Control* **2018**, *91*, 100–112. [[CrossRef](#)]
24. Laine, A.; Högnäsbacka, M.; Niskanen, M.; Ohralahti, K.; Jauhiainen, L.; Kaseva, J.; Nikander, H. Virallisten Lajikekoikeiden Tulokset 2009–2016 (Results of the Official Variety Trials 2009–2016). Luonnonvara- ja Biotalous Tutkimus 1/2017, 271p. Natural Resources Institute Finland (Luke). 2017. Available online: <http://urn.fi/URN:ISBN:978-952-326-346-8> (accessed on 15 June 2024).
25. Saari, E.E.; Prescott, M. A scale for appraising the foliar intensity of wheat diseases. *Plant Dis. Rep.* **1975**, *59*, 377–379.
26. Finnish Meteorological Institute. Open Database. Available online: <https://www.ilmatieteenlaitos.fi/havaintojen-lataus> (accessed on 5 September 2023).
27. Dash, M.; Liu, H. Feature selection for classification. *Intell. Data Anal.* **1997**, *1*, 131–156. [[CrossRef](#)]
28. Blum, A.L.; Langley, P. Selection of relevant features and examples in machine learning. *Artif. Intell.* **1997**, *97*, 245–271. [[CrossRef](#)]
29. 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](#)]
30. Uncu, Ö.; Türkşen, I.B. A novel feature selection approach: Combining feature wrappers and filters. *Inf. Sci.* **2007**, *177*, 449–466. [[CrossRef](#)]
31. 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](#)]
32. Ruusunen, M. Signal Correlations in Biomass Combustion—An Information Theoretic Analysis. Ph.D. Thesis, University of Oulu, Oulu, Finland, 2013.
33. Fisher, R.A. The Use of Multiple Measurements in Taxonomic Problems. *Ann. Eugen.* **1936**, *7*, 179–188. [[CrossRef](#)]
34. Theodoridis, S. Chapter 7—Classification: A Tour of the Classics. In *Machine Learning*, 2nd ed.; Elsevier: Amsterdam, The Netherlands, 2020; pp. 301–350. [[CrossRef](#)]
35. The MathWorks, Inc. Statistics and Machine Learning Toolbox Documentation. Available online: <https://www.mathworks.com/help/stats/index.html> (accessed on 7 February 2023).
36. Grandini, M.; Bagli, E.; Visani, G. Metrics for Multi-Class Classification: An Overview. *arXiv* **2020**. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.