
MECHANIZATION, ELECTRIFICATION,
AUTOMATION AND DIGITALIZATION

AI Solutions for Digital Diagnostics of Grain Crop Diseases (Based on the Example of *Pyrenophora teres* in Winter Barley)

I. V. Arinichev^{a, *}, I. V. Arinicheva^b, G. V. Volkova^c, and Y. V. Yakhnik^c

^a Kuban State University, Krasnodar, 350040 Russia

^b Trubilin Kuban State Agrarian University, Krasnodar, 350044 Russia

^c Federal Scientific Center for Biological Plant Protection, Krasnodar, 350039 Russia

*e-mail: galvol.bpp@yandex.ru

Received November 10, 2023; revised December 19, 2023; accepted January 9, 2024

Abstract—The purpose of this research is to justify the feasibility of using digital intelligent technologies in forecasting the development of net blotch in winter barley. The developed AI solution is a binary decision tree that can predict scenarios of net blotch development: depressive, moderate, and epiphytotic development. To configure the algorithm parameters, we carried out field and laboratory experiments at the Federal Scientific Center for Biological Plant Protection from 2021 to 2023. The preparation of data involved several stages, including setting up of field plots to create an artificial infection background as well as the preparation of an inoculum, sowing of highly susceptible and resistant winter barley varieties, and artificial inoculation. The selected input factors included the observed degree of leaf damage, type of variety resistance, vegetation phase at the time of primary infection, and average relative air humidity during the vegetation phase of infection. The total sample size was 144 observations. The trained model has demonstrated a high classification accuracy on both the training and test datasets at an accuracy rate of more than 96%. Based on the statistical estimate of the significance of the factors influencing the development of net blotch in barley, it is shown that the most influential factor is the current degree of leaf infection (74.3%), followed by the average relative air humidity (11.9%), the resistance of the variety to the disease (10.4%), and the development stage during which infection occurred (3.4%). The proposed solution has a significant practical importance since it provides new opportunities for the diagnostic process of net blotch in winter barley, including high diagnostic rate, accuracy in forecast predictions, and applicability in field conditions.

Keywords: grain production, winter barley (*Hordeum vulgare* L.), net blotch (*Pyrenophora teres*), digital diagnostics, artificial intelligence

DOI: 10.3103/S1068367424700034

INTRODUCTION

Grain production is the largest subsector of agriculture; its stable development plays a critical role in achieving food security in the country and is strategically important in solving a number of problems, including the supply with safe, high-quality, and affordable agricultural products for the country's population, the creation of jobs in rural areas, the development of agricultural infrastructure, strengthening of the country's economic position in world markets, etc. Grain crops, such as wheat, barley, rice, and corn, provide a steady flow of products for both domestic consumption and export and serve as the foundation of the world food market.

Despite the steady increase in the gross grain harvest over the past 5 years (from 113.3 million t in 2018 to 157.7 million t in 2022) [1] and the confident crossing of the established threshold value of the proportion of domestically produced grain (95% according to the Food Security Doctrine of the Russian Federation)

(in 2022, the level of self-sufficiency exceeded 150%) [2], the industry faces new and increasingly serious challenges, the most significant of which is threat to the phytosanitary state of crops.

Among the main factors of mass spread of diseases and pests in grain crops, the following factors are decisive:

(1) Deterioration of weather and climate conditions, including an increase in average annual temperatures in most grain-producing regions, unexpected and prolonged precipitation in later growing seasons, and extreme weather events (droughts, floods, and abnormal temperatures), which contribute to the spread of diseases and pests.

(2) Frequent and untimely use of chemical plant protection products, which leads to the development of resistance of pathogens to agents used, makes conventional control methods less effective, and requires more integrated approaches.

(3) Monoculture, which promotes the accumulation of pathogens and pests in soil, thereby increasing

the risk of disease spread and aggravating phytosanitary problems.

(4) Insufficient attention to the system of monitoring activities that are performed as part of production activities and crucial for detecting, predicting, and monitoring the phytosanitary state of crops.

There is no doubt that it is impossible to solve one of the main problems in grain production (mass spread of diseases and pests) without transition to innovative methods of production. Today, artificial intelligence is coming to the forefront of all innovative solutions; it determines the core of such innovations and has enormous potential for digital transformation.

The development of intelligent solutions opens up new opportunities for increasing sustainability, reducing the dependence, minimizing manual labor, and, consequently, increasing the efficiency of the entire industry, which helps solve complex challenges and ensure the food security of the economy, region, and whole country [3].

An intelligent diagnosis of net blotch of barley caused by the fungus *Pyrenophora teres* Drechsler, one of the dominant pathogens in the cultural cenosis both in the south of Russia and throughout the world, was previously carried out. Annual crop losses due to its impact range from 15 to 50% and the frequency of epiphytosis is five times every 10 years. Under the favorable combination of factors (weather, variety susceptibility, and vegetation phase), the prevalence of the disease can reach 100% and develop to 50–90% [4, 5].

The relevance of digital diagnostics of net blotch is determined by the low effectiveness of classical methods of combating this disease, including agrotechnological methods, seed treatment, and cultivation of resistant varieties. The reliable preservation of crops can be achieved only using effective fungicides. In this case, the control point in the production process is making a decision on the advisability of their use in a specific period of time.

To support these decisions, models were previously developed based on artificial intelligence (AI models), which made it possible to solve problems of detection and classification of *Pyrenophora teres* with respect to other types of diseases with similar symptoms [6] as well as problems of localization of affected areas and determination of the degree of damage to leaves [7], which serves as one of the main signals for using chemical agents.

However, along with the above-listed tasks, where AI solutions have already shown good results both as a result of the reduction in diagnostic time and as a result of an increase in the proportion of accurate predictions, issues related to the prediction of disease development are of high practical value and relevance. It is important not only to diagnose the disease and state the degree of damage to plants at the current moment but also to forecast the disease progression.

The purpose of this research is to justify the feasibility of using digital intelligent technologies in the prediction of the development of net blotch in winter barley.

MATERIALS AND METHODS

To achieve the set goal, we carried out field and laboratory studies at the sites of the Federal Scientific Center for Biological Plant Protection in 2021–2023, which involved the infection of plants with a population of *Pyrenophora teres*, and monitored primary manifestations and development dynamics. The research involved three winter barley varieties (Vivat, Rubezh, and Romans), which are sown in the south of Russia and differ in the resistance to net blotch pathogens (resistant and susceptible varieties). During the experiments, we used classical phytopathological methods and approaches. *P. teres* was identified using the key of V.I. Bilai [8]. The production of fungal inoculum and inoculation in the full tillering phase under field conditions were based on standard methods [9]. Records were carried out starting from the primary manifestation of the disease to the phase of milky-waxy grain ripeness at an interval of 10–12 days. The degree of damage to leaves and other organs by net blotch was determined according to the Heschele scale.

As input model factors, which are recorded during the implementation of the experiment and influence the degree of net blotch development, we used the observed degree of leaf damage (%), type of variety resistance (R = resistant and S = susceptible), vegetation phase at the time of primary infection (tillering, booting, and flag leaf phase), and average relative air humidity in the vegetation phase at the time of infection (%).

The output target variable is the degree of development of the disease in the phase of early milky ripeness with the following possible values: D, depression, M, moderate development, and E, epiphytosis. The class marks of the resulting variable were selected based on [10], the authors of which predicted the phytosanitary state of wheat crops.

The total sample size was 144 observations corresponding to different combinations of the desired features. Among them, 115 objects were randomly used as a training set and 29 objects for the final assessment of the model quality. Splitting a sample into a training set and a test set is an important stage in the methodology of machine learning and applied statistics; its purpose is to assess the performance of the model and its generalization ability. The main idea behind this splitting is to “hide” part of the data from the algorithm that are used during the configuration of the model parameters to be used for checking the quality at the necessary stage. If the model can make accurate predictions on the test part of the data that were not previously input, its use makes it possible to generalize and apply the knowledge gained during the training process to new data.

During the analysis of the applicability of different models for predicting the development of net blotch in winter wheat, the family of decision tree algorithms was chosen as a basic family. The final model determining sample objects for different classes of net blotch development was trained according to generally

accepted machine-learning methods (stochastic gradient descent, error backpropagation, etc.).

RESULTS AND DISCUSSION

The choice of decision trees for predicting the development of net blotch in winter barley is justified by several key reasons:

(1) Tree algorithms provide clear and interpretable results. This is especially important for agriculture, where food producers should make forecasts without using special software. With this algorithm, it is sufficient to move from the root vertex to one of the end vertices, check the correspondence of the conditions at the model nodes, and forecast the disease development.

(2) The use of a decision tree makes it possible to relatively quickly process both small and huge volumes of data, which ultimately makes it possible to quickly perform a predictive function.

(3) Decision trees can be scaled to new conditions and task requirements, which makes them suitable for a variety of situations, in particular, for other crops and their pathogens.

(4) Models based on decision trees can work with categorical features that are not expressed on a numerical scale, which is especially valuable for agricultural production, where qualitative factors often serve as predictors. In our case, two of the four factors were categorical.

The decision tree was constructed using the recursive machine-learning procedure, *Induction of Decision 3* (ID3). Its most important step is searching for a predicate (binary function) that is assigned to each internal vertex of the tree and splits the sample into two parts. It comes down to optimizing the branching information criterion, a special metric that helps the algorithm determine the feature that is best suitable to separate data at each tree level.

This study used the most common Gini index, which shows how many pairs of objects belonging to the same class will simultaneously fall into the left or

right child node of the tree (the predicate values coincide on these pairs):

$$I(\beta, X') = \#\{(x_i, x_j) : y_i = y_j, \beta(x_i) = \beta(x_j)\},$$

where X' is the training set, x_i and x_j are the features (factors) of objects in the training set, β is the predicate, and $\#$ is the operator that counts the number of pairs meeting the condition.

The net blotch prediction model was built in *Python*. To implement the decision tree algorithm, determine the model parameters, and draw the model (Fig. 1), we used the *scikit-learn 1.0.2* library, which provides extensive functionality for machine learning. Each block of the constructed tree has the same type of structure:

(1) The name of the factor is the feature by which the sample is split into two subgroups. The tree node compares the value of this feature with a certain threshold and the data are sent to one of the tree branches, depending on the result.

(2) The Gini index (“gini”) measures the degree of “confusion” of classes in the node. The lower its value, the “purer” the node.

(3) Sample size (samples) is the number of observations (specimens) in the tree node.

(4) The value contains information about the distribution of classes in the node. For example, if the node splits the sample into two subgroups, the value would be the number of specimens of each class in each subgroup.

(5) Class: if the node is a leaf (end) node (i.e., if it has no child nodes), this parameter reflects the predicted class for this leaf.

Each internal vertex of the tree reflects the feature by which the sample is split into two sets. “Rapid coding” was previously carried out for all qualitative features (*one-hot encoding*), which represents categorical features as binary vectors, thereby making them suitable for use in machine-learning models, including decision trees. Thus, ordinary binary coding was performed for the “variety resistance” factor: 0 is resistant variety and 1 is susceptible variety. For the “vegetation phase” factor, coding was carried out using two binary variables:

$$\text{Tillering_phase} = \begin{cases} 0, & \text{observation does not cover the tillering phase,} \\ 1, & \text{observation covers the tillering phase.} \end{cases}$$

$$\text{Flag_leaf} = \begin{cases} 0, & \text{observation does not cover the flag leaf phase,} \\ 1, & \text{observation covers the flag leaf phase.} \end{cases}$$

After coding, three possible vegetation phases, which were recorded in the experiment, were presented in the form of paired combinations of values of dummy variables ((1, 0) tillering phase; (0, 1) flag leaf phase; and (0, 0) booting phase) and then sent to the input of the model.

Marks contained in the terminal (leaf) vertices were the classification result. For assessing the quality

of the algorithm implementation on the training set, it is sufficient to compare the distribution of the true values of the target variable for all objects included in the leaf vertex (*value* vector) with the value of the predicted class. For instance, if the *value* vector is [0, 13, 0] in the leaf node and the model forecast produces “Moderate development,” it can be stated that all the 13 observations were classified correctly. When the

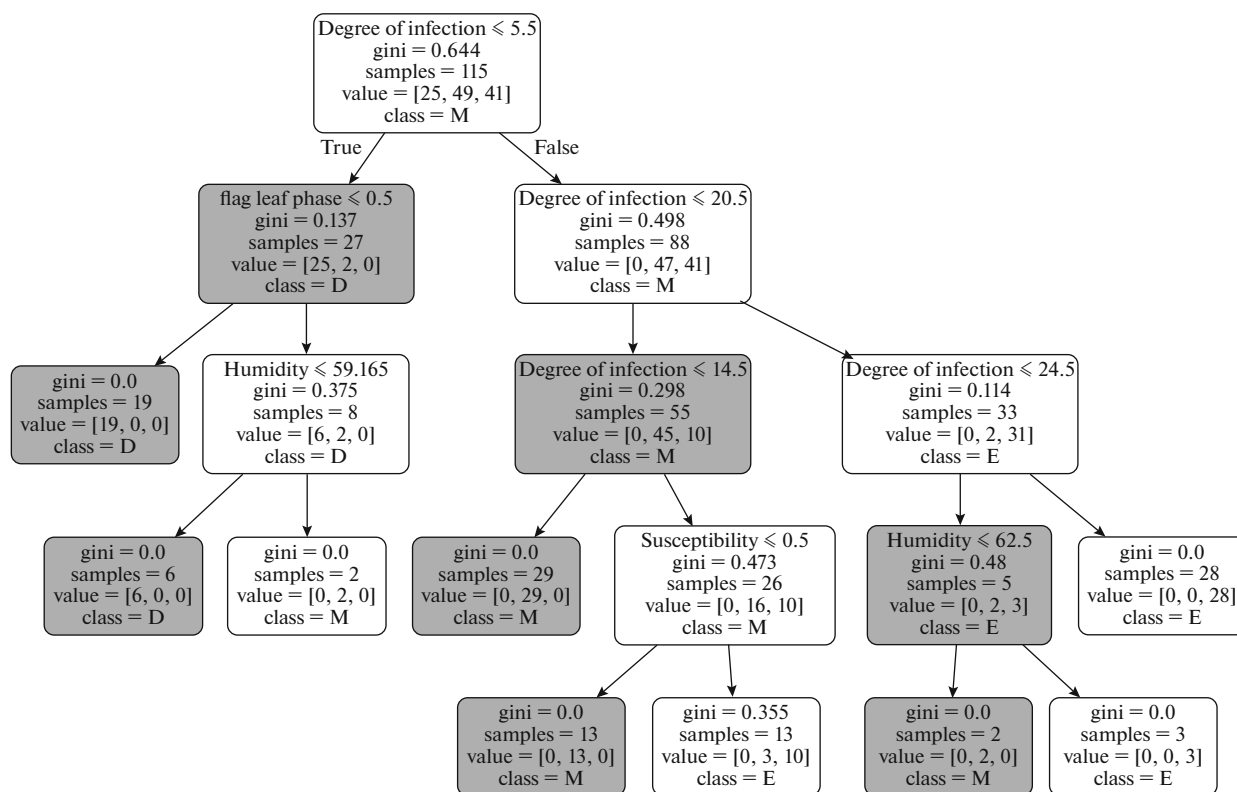


Fig. 1. Trained decision tree predicting the development of net blotch in winter barley.

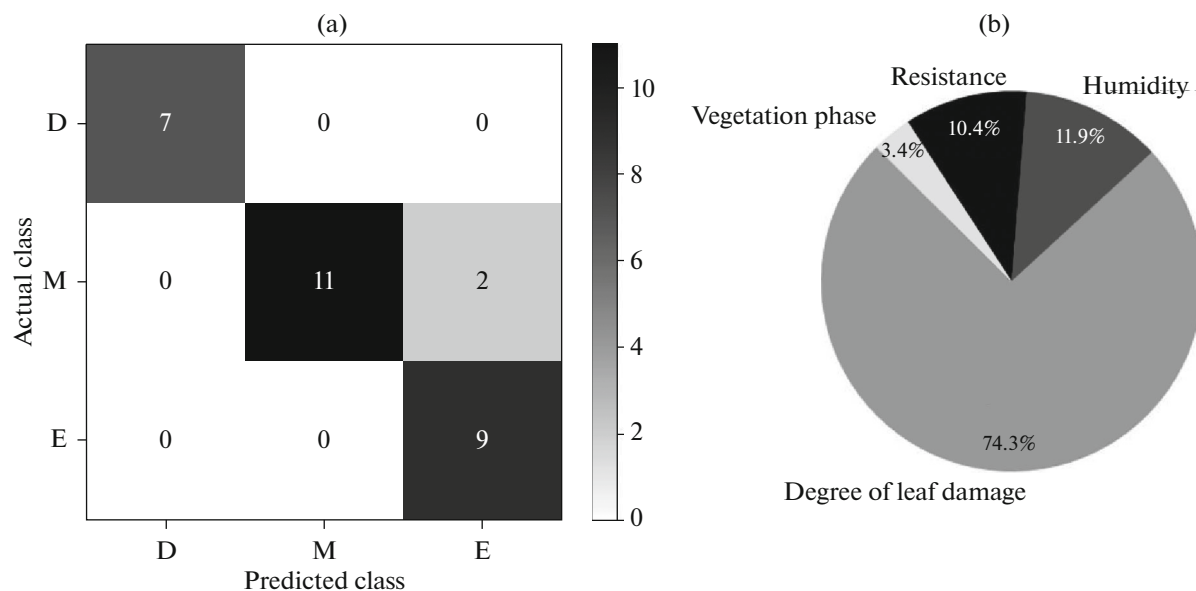


Fig. 2. Modelling results: (a) algorithm error matrix; (b) significance of factors in modeling.

value is [0, 3, 10] and the model predicts “Epiphytotic development,” the algorithm made an error three times by classifying three observations with the moderate development of net blotch as the most pessimistic scenario. In our studies (see Fig. 1), this case of

misclassification was detected in one single end vertex. The percentage of correct answers of the algorithm (*accuracy*) on the training set was 98.2%.

Naturally, assessment of the quality of modeling using the training set alone may be insufficiently

informative. This is determined by the phenomenon called *overfitting* (when the model overfits the training data and is poorly effective on new test sets).

A detailed examination of the error matrix of the algorithm on the test set (Fig. 2a) suggests that the trained model correctly classified most of the objects based on new data and made an error on just one observation. This indicates its ability to generalize data, the absence of overfitting, and its suitability for practical use.

Information about the significance (weights) of factors (Fig. 2b) used by the model during decision making shows the extent to which each feature influences the forecast. The development of net blotch in barley is most significantly influenced by leaf damage (74.3%). The proportions of variety resistance and relative humidity were at the level of 10% and the “vegetation phase” of barley proved to be the least significant feature.

The presented results have a number of important practical and scientific applications:

(1) With account for the scaling feature of decision trees and the collected data on other crops and diseases, the presented AI solutions may become a tool for making management decisions since they provide a quick and objective assessment of the degree of their development.

(2) The use of the model contributes to a more efficient distribution of time and material resources, which is ultimately reflected in the efficiency of production activities (owing to the reduction of production costs).

CONCLUSIONS

Based on modern machine-learning methods, an AI model was developed for predicting the development of net blotch in winter barley. A decision tree was chosen as the basic algorithm. The model built according to the training set demonstrated a high predictive capacity on test data: the proportion of correct answers was 98% on the training set and 97% on the delayed set.

The main factors influencing the development of net blotch in barley are the current degree of leaf infection (the contribution is 74.3%), average relative air humidity (11.9%), variety resistance to the disease (10.4%), and the vegetation phase at the time of infection (3.4%).

FUNDING

This study was supported by the Kuban Science Foundation as part of scientific project no. MFI-20.1/121. No additional grants were received to conduct or direct this specific research.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This work does not contain any studies involving human and animal subjects.

CONFLICT OF INTEREST

The authors of this work declare that they have no conflicts of interest.

REFERENCES

1. Federal State Statistics Service. Agriculture, hunting and forestry. https://rosstat.gov.ru/enterprise_economy. Cited October 1, 2023.
2. Decree of the President of the Russian Federation dated January 21, 2020 No. 20 “On approval of the Doctrine of Food Security of the Russian Federation.” <http://www.kremlin.ru/acts/bank/45106>. Cited October 1, 2023.
3. Skvortsov, E.A., Nabokov, V.I., Nekrasov, K.V., et al., Application of technologies of artificial intelligence in agriculture, *Agrar. Vestn. Urala*, 2019, vol. 8, no. 187, pp. 91–98. https://doi.org/10.32417/article_5d908ed78f7fc7.89378141
4. Volkova, G.V. and Astapchuk, I.L., The spread of *Pyrenophora teres* on barley crops in the North-Caucasus region, *Zernovoe Khoz. Ross.*, 2019, vol. 2, no. 62, pp. 63–68. <https://doi.org/10.31367/2079-8725-2019-62-2-63-68>
5. Orlovskaya, E.N., Astapchuk, I.L., and Volkova, G.V., Influence of soil tillage system on sensitivity of winter barley to barley net blotch, *Vestn. Ul'yanovsk. Gos. S-kh. Akad.*, 2018, vol. 4, no. 44, pp. 110–177. <https://doi.org/10.18286/1816-4501-2018-4-110-117>
6. Arinicheva, I.V., Arinichev, I.V., and Repko, N.V., Automated diagnostics of barley fungal diseases, *Tr. Kuban. Gos. Agrar. Univ.*, 2022, no. 95, pp. 69–72. <https://doi.org/10.21515/1999-1703-95-69-72>
7. Arinichev, I.V., Polyanskikh, S.V., and Arinicheva, I.V., Semantic segmentation of rusts and spots of wheat, *Komp'yuternaya Optika*, 2023, vol. 47, no. 1, pp. 118–125. <https://doi.org/10.18287/2412-6179-CO-1130>
8. Bilai, V.I., *Mikroorganizmy – vzbuditeli boleznei rastenii (spravochnik)* (Microorganisms – Pathogens of Plant Diseases (Handbook)), Kiev: Naukova Dumka, 1988.
9. Khasanov, B.A., *Opredelitel' gribov-vzbuditelei "gel'mintosporiozov" rastenii iz rodov Bipolaris, Drechslera i Exserohilum* (Guide to Fungal Causal Agents of “Leaf Spot Diseases” of the Genera Bipolaris, Drechslera and Exserohilum), Tashkent: Fan, 1992.
10. Ibragimov, T.Z. and Sanin, S.S., Digital plant protection and phytosanitary information data mining, *Zashch. Karant. Rast.*, 2019, no. 4, pp. 15–18.

Translated by D. Zabolotny

Publisher’s Note. Allerton Press remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.