

Predictive Modeling of Net Blotch Disease in Spring Barley Using Artificial Neural Networks (ANN) and XGBoost: A Comparative Study

Recep Sinan ARSLAN*¹, Nilüfer AKCİ², Dilara ÇELİK³, Bahatdin Daşbaşı⁴, Kadir Aytaç ÖZAYDIN⁴, Teslima DAŞBAŞI⁵

*1 Kayseri University, Department of Computer Engineering, Kayseri, Türkiye

2 Plant Production Central Research Institute, Ankara, Türkiye

3 Kayseri University, Department of Computer Engineering, Kayseri, Türkiye

4 Kayseri University, Department of Basic Engineering Sciences, Kayseri, Türkiye

5 Central Research Institute, Ankara, Türkiye

6 Kayseri University, Bünyan Vocational School, Kayseri, Türkiye

(Alınış / Received: 08.04.2025, Kabul / Accepted: 28.05.2025, Online Yayınlanma / Published Online: 30.08.2025)

Keywords

Barley,
Net Blotch,
Machine Learning,
Artificial Neural Network,
Advanced Feature Analysis.

Abstract: The detection of plant diseases and their application to agricultural production have the potential to increase productivity and regulate pesticide use. Time-dependent meteorological climate data play an active role in plant disease control practices. The aim of this study was to develop a prediction model for the presence of net blotch disease in barley plants in Turkey. The machine learning and artificial neural network-based model was developed and evaluated using a three-year disease development period (2021-2023) as the training and testing dataset. The model was trained using nine different meteorological climate data, including temperature (min, max, avg), humidity, rainfall, wind speed, sun exposure time, actual vapor pressure, dew point temperature. The outcomes of the evaluation process yielded an 88.9% accuracy for the XGBoost classifier and a noteworthy 92.16% accuracy value for the Artificial Neural Network (ANN). The precision, recall and f-score values for the ANN model were found to be 96.10%, 89.16% and 92.50%, respectively. The findings indicated that the model exhibited superior performance in detecting the absence of the disease when compared to existing studies. Furthermore, the impact of nine distinct inputs on the manifestation of the disease and the interrelationships between these inputs were analysed. The analysis revealed that sun exposure time, temperature, wind speed and humidity factors exerted the most substantial influence on environmental variables. The high performance obtained from the model lends further credence to the hypothesis that artificial intelligence models can be successful in combating the disease and that an increase in productivity can be achieved in the product by reducing the negative effects of the disease.

Yapay Sinir Ağları (YSA) ve XGBoost Kullanılarak İlkbahar Arpasında Net Leke Hastalığı Tahmin Modeli: Karşılaştırmalı Bir Çalışma

Anahtar Kelimeler

Arpa,
Net Leke,
Makine Öğrenmesi
Yapay Sinir Ağı
Gelişmiş Özellik Analizi

Özet: Bitki hastalıklarının tespiti ve bunun tarımsal üretime uygulanmasıyla üründe hem verimlilik artışının sağlanması hemde pestisit kullanımının düzenlenmesi mümkün olabilecektir. Bitki hastalıkları ile yapılacak mücadele uygulamalarında zamana bağlı meteorolojik iklim verileri etkin rol oynamaktadır. Bu çalışma kapsamında Türkiye’de arpa bitkisinde net blotch hastalığının varlığına ilişkin bir tahmin modelinin geliştirilmesi amaçlanmıştır. Sınıflandırma için sıcaklık, nem, yağış miktarı, rüzgar hızı, güneşlenme şiddeti, gerçek buhar basıncı ve çiğ noktası

sıcaklığı olarak verilen 9 farklı meteorolojik iklim verilerinin makine öğrenmesi ve yapay sinir ağı tabanlı modelin eğitim ve test süreçlerinde kullanılması yaklaşımı benimsenmiştir. 2021-2023 yılları arasındaki 3 yıllık bir hastalık gelişimi dönemine ait veriler kullanılmıştır. Yapılan testler sonucunda XGBoost sınıflandırıcı ile 88.9% Accuracy değeri elde edilirken, Yapay sinir ağı (YSA) ile 92.16% Accuracy değeri elde edilmiştir. YSA modeli için precision, recall ve f-score değerleri sırasıyla 96.10%, 89.16% ve 92.50% olmuştur. Sonuçlar modelin hastalığın yokluğunu tespit etmede literatürdeki çalışmalara kıyasla en iyi başarıya sahip çalışmalarından birisi olduğunu göstermiştir. Bunun yanında 9 farklı girdinin hastalığın meydana gelmesine etkileri ve girdilerin kendi aralarındaki ilişkileri de analiz edilmiştir. Sonuçta, sıcaklık, güneşlenme süresi ve nem faktörlerinin, çevresel değişkenler üzerinde en belirgin etkiye sahip olduğu anlaşılmıştır. Elde edilen yüksek başarımlar, hastalığın mücadelesinde yapay zeka modellerinin başarılı olabileceğini ve hastalığın olumsuz etkisini düşürerek üründe verimlilik artışının yakalanabileceğini kanıtlamıştır.

1. Introduction

Plants are living beings that can produce food through photosynthesis and have roots, stems and leaves. There are many plant species around the world, and these species range from microscopic algae to plants that can grow up to 90 meters tall. Ambient temperature is an important factor affecting plant development. High temperatures increase the plant's water loss, while low temperatures cause water to freeze. Light also affects plant development, and plants only use light at certain wavelengths. In addition to other factors, gases also play a role in plant development; oxygen and carbon dioxide affect the balance of respiration and photosynthesis, while sulfur dioxide and ethylene change the activities of hormones. Some plants develop and secrete defense mechanisms to protect themselves from external factors. Plants play an important role in maintaining ecological balance and sustaining life on earth. They also prevent erosion and are indispensable resources for humanity in many areas such as agriculture, food and industry [1].

Cereals are starch-rich grains belonging to the Poaceae family in agricultural activities. These grains are grown for their grains and are an important source of nutrition for both human and animal nutrition. Cereal types include wheat, barley, rye, corn, rice, oats and millet [2]. The most cultivated and produced plant group in the world is cereals. In addition, Turkey has an important place in the world's cereal production [3]. Provinces such as Konya, Şanlıurfa, Ankara, and Diyarbakır are among the provinces with the largest grain production. According to different climate conditions, grain types are divided into cool climate grains such as wheat, barley, and oats, and hot climate grains such as corn and rice. Grain cultivation in Turkey covers a large part of our agricultural lands and forms the basis of the country's plant and animal production [4]. In rural areas, it is produced for nutrition, feed needs and as a source of income.

Among cereals, barley is the fourth most produced plant in the world after corn, wheat and rice [5]. Barley, one of the cultivated plants used in human nutrition, is at the forefront in terms of production amount as well as cultivation area both in the world and in our country thanks to its wide adaptability.

Barley is a strategic product in terms of being a raw material for agricultural industry in the world and in Turkey, contributing to the economy and being the main source of income for rural areas. Barley, which can be grown in every region in Turkey, ranks second after wheat in terms of cultivation area and production amount among field products. Central Anatolia and Southeastern Anatolia Regions in particular are two important regions in barley production. According to 2023 statistics, barley production increased by 8.2% compared to the previous year and reached 9.2 million tons [6]. Barley (*Hordeum vulgare* L.), which is grown both for animal feed and as beverage malt, is one of the important grain sources in the world. Its production in the world between 2022-2023 was 150.9 million tons and Turkey ranks 5th in the world in barley production [7].

Global warming refers to the process of increasing average temperatures worldwide due to the increase in greenhouse gases in the atmosphere. These gases consist of compounds such as CO₂, CH₄, N₂O, O₃ and CFC in the atmosphere and keep some of the heat coming from the sun, allowing the earth to remain at certain temperatures. This effect is called the greenhouse effect and is necessary for the sustainability of life, but excessive greenhouse gas emissions lead to global warming. In this case, it leads to increasing temperatures worldwide, climate change, increased frequency of extreme weather events, and destruction of ecosystems [8, 9]. In the direct effects of global warming, in addition to increasing temperatures, it also negatively affects crop productivity. This creates major risks for food security [10]. Climate change threatens biodiversity by causing serious damage to ecosystems [11].

In addition, the effects of global warming pose serious threats to the use of water resources. These changes may lead to a decrease in water resources and an increase in water-related problems [12]. Global warming and climate change have major effects, especially on agriculture. Agricultural activities are very negatively affected by industrial activities and environmental destruction. The increase in industrial activities accelerates global warming by increasing environmental pollution and has direct effects on agricultural activities [13]. For these reasons, a global effort is necessary to combat global warming, reduce greenhouse gas emissions and protect the environment.

These environmental conditions that change with global warming significantly affect the formation, severity, prevalence, geographical distribution of plant diseases and their control strategies. In addition, climate change can change the period and density of the pathogen, the resistance status of the host and there may be changes in the geographical distribution of pathogens, therefore it shows that changes in control strategies are necessary [14]. In addition, one of the biggest obstacles to increasing production and quality in cereal farming both in the world and in our country is diseases, which are biotic stress factors. In the event of an epidemic, diseases can negatively affect product yield and quality and cause economic losses at different levels. Due to climate change, products such as wheat, barley, rice, sugar beet, potato and corn, which are the main agricultural products in tropical and temperate regions, are negatively affected. Plant development and yield are significantly affected by high carbon dioxide concentration, temperature, changes in precipitation regime and extreme weather conditions. The course of plant diseases also changes under these changing conditions. In the case of pathogens with short life cycles, their reproductive capabilities and disease spreading capacity will increase and adaptation to climate change will occur rapidly. In addition, climate change; human-induced reasons such as pollution of air, water and soil, transportation of exotic species to long distances and factors such as urbanization, plant diseases will be affected more. These factors will contribute to the rapid spread of diseases [15].

In order to prevent barley diseases and prevent the disease from becoming an epidemic, it is important to actively use modeling methods and develop early warning and prediction systems. Thus, with the studies to be carried out, it is to help farmers, decision makers and publishers with strategic support. In addition, due to changing climate conditions, modeling studies are needed against diseases that are common in cereal groups, which are the homeland and genetic resource center of our country.

The causative agent of the disease is *Pyrenophora teres* Drechs, a fungus which has two different forms. The first, net blotch, is called *Pyrenophora teres* f. *teres* and consists of thin, dark brown, longitudinal lines on the leaves. The second, the spot form, *Pyrenophora teres* f. *maculata*, contains small necrotic spots surrounded by chlorosis. Since 2009, these diseases have been observed in the Marmara, Aegean, Central Anatolia, Black Sea and Southeastern Anatolia regions, with net blotch disease causing yield losses of up to 40% in years with heavy rainfall. The prevalence of net blotch disease has been documented in 80.2% of the Thrace region and 51.2% of the Aegean region. The study also determined that the average rate of disease-infected areas in Central Anatolia is 14.3% and 12.9% in Southern Marmara. Consequently, it was concluded that the most conducive conditions for the development of the disease are in Thrace, as well as in coastal and transitional regions. The analysis of five-year research data indicates that approximately half of the barley planting areas are infected with barley net blotch disease, which results in an annual economic loss of 23% in the overall yield [16].

Cultural measures are the primary methods used to combat barley net blotch disease. Cultural measures are measures to be taken to prevent the disease or reduce its damage. As a cultural measure; Destroying diseased plant residues and using disease-free seeds are the primary ones. In addition, cultural measures such as fertilizing at appropriate rates and times should be taken to ensure the strong development of the barley plant. Among the methods used to combat the disease, developing resistant varieties is important. Using resistant barley varieties and planting resistant varieties are very effective in combating the disease [17]. Chemical control of barley net blotch disease is carried out in the form of green part application. It should be done when the spots of the disease agent are first seen and in a way that covers all green parts. In these chemical control applications, it is important to follow the disease quickly, take observations and start the control. The disease is provided with conidia emergence within 48 hours in a humid environment at 10-18°C. The mycelia in the seed coat and pericarp of infected seeds can be the first source of infection. A 24-hour dark and humid period is sufficient for the disease to infect the leaves. Chemical control applications should be carried out by producers by taking into account the climatic conditions from the moment the disease is first seen. Thus, with these applications, yield and quality losses in the product are reduced and the desired level of product can be obtained.

There are not many studies in the literature on the prediction of net blotch disease in barley. The number of studies on this subject is limited and the summary is presented below.

Fenu et al. [18], prepared a review containing studies aiming to use big data analysis to find a solution to the huge losses in food products due to plant diseases every year. In this review, the analysis of the effects of environmental factors affecting the disease and the studies conducted for the early detection of diseases by using these effects were examined. The studies included tests of RF, SVM, Regression analysis, ANN, LSTM and statistical models. They stated that CNN-based models have low RMSE and MAE values in terms of models using long-term data. When considered as a classification problem, they achieved the highest classification performance with Least Squares SVM.

Charaya et al. [19], produced an ANN model using different input values such as host factors (susceptibility/resistance, density, distribution), plant growth environment, pathogen factors, environmental factors for the prediction of any plant disease. Thus, they aimed to predict disease severity with this method. In their study, algorithmic range values were found and shared for tomato, rice, barley and potato. In this direction, they aimed to analyze the values for which prediction is possible and which affect the prediction.

Landschoot et al. [20], developed a model to predict the incidence of Fusarium head blight disease in wheat in a study based on 3100 wheat samples collected from 18 regions in Belgium between 2002-2021. The model, with a web interface, allows predictions to be made specific to the relevant field. The weather forecast and agronomic variables closest to the field were used for the prediction. In this way, they enabled the development of strategies for the prediction and management of the disease. Musa et al. [21], similarly analyzed Fusarium head blight disease in terms of its effect on grain production in Switzerland. By examining wheat planting areas, the effect of the disease on wheat was observed and they developed a decision support system called FusaProg. Unlike other studies, the model they proposed takes into account planting factors, characteristics of the product planted in the previous year and soil characteristics instead of meteorological data. In this model, it estimates the infection risk periodically with 3 different threshold-based algorithms and suggests chemical control application for this estimate. Shah et al. [22] investigated the effects of weather conditions on plant diseases. They analyzed the flowering period of Fusarium head blight disease in wheat for 140 days and its relationship with climate parameters. As a result of the modeling they made with the logistic regression model, an average precision value of 0.75 was reached and the recall value was at a similar level. Shah et al. [23], again with modeling studies using logistic regression, made the estimation and time-dependent analysis of Fusarium head blight disease severity in the USA. In this study, where rainfall amounts were used in estimation, it was evaluated that the effects of relative humidity or temperature were less than the precipitation-dependent estimator. In another study conducted in the USA, Shah et al. [24] analyzed the effect of climate parameters in the prediction of Fusarium head blight disease. 380 different variables representing temperature, relative humidity and precipitation were used in the logistic regression model. The results showed that relative humidity characterized the formation and development of the disease more than other factors. The values obtained showed 19% better performance than the existing systems.

Jorgensen et al. [25], investigated the necessity of disease risk models to prevent unnecessary fungicide use to prevent diseases. They conducted studies in 33 barley fields grown in 5 countries and during 2 production seasons in the Scandinavian-Baltic region to verify the favorable conditions for the formation of net blotch disease in barley. The number of rainy days, resistance of varieties and disease development data were used as inputs. In the study, it was determined that 50-60% for a single fungicide application and 69-80% for two fungicide applications and it was verified by comparing with the yield increase in the product. They stated that the model aimed to reduce fungicide use with risk analysis for disease formation and thus to provide yield increase.

Henriksen et al. [26], proposed a decision support system called PC-P for cereal disease management. The proposed model aims to determine the doses for combating barley diseases and to increase the yield in the product accordingly. The model analyzes the input parameters algorithmically according to their ranges and makes decisions on determining the drug application doses. In the tests conducted for barley diseases, it was shown that the highest product yield was achieved by reducing the appropriate chemical control dose by 50% for the prevention of the disease.

Ruusunen et al. [27], conducted a study in Finland where they assessed the risk of plant diseases using meteorological climate data. Data including plant disease observations between 1991-2007 were obtained from the Meteorological Institute. They used features including mean temperature, minimum temperature and precipitation to determine the probability of net blotch disease causing an epidemic in barley. They estimated the epidemic status of the disease to be 3 levels. Disease prediction was made sequentially in 14-day periods. In the same researchers next paper [28] aimed to find the optimum onset time of net blotch disease in barley using meteorological data. As in their previous studies, they provided analysis of the features using a temporal period and mathematical transformations. They determined that examining in periods of 41-55 days on average is more useful in terms of determining the disease period. They showed that the disease can be determined with a minimum confidence interval of 68% in a 9-day period within a 14-day period. In their latest study conducted in

2024, Ruusunen et al. [29] analyzed the classification of the epidemic status of barley net blotch disease into 3 levels. For the classification, they adopted the method of mathematically converting meteorological climate data, dew point temperature, humidity and outdoor temperature values. In tests conducted with 36 years of data, it was shown that net blotch disease in Finland was determined with high success in 14-day periods. Thus, it was aimed to reduce the use of pesticides for the disease.

When the studies in the literature are examined, the main purpose is to reduce pesticide use and to help decide when the disease can be seen [19, 18]. The risks assessed in these studies are generally based on pathogen and host (plant) factors, selected meteorological climate parameters, in some cases, data on the occurrence and symptom formation of diseases in previous periods, and geological location [25, 26]. Prediction systems for the determination of Fusarium head blight disease, one of the other fungal diseases, have been examined in some studies. Prediction systems for the determination of Fusarium head blight disease, one of the other fungal diseases, have been examined in some studies [20, 21, 22, 24]. A study on the development of a model for predicting the epidemic status of barley net blotch disease was conducted in Finland [27, 28]. Our main difference from these studies is that it is one of the first studies conducted according to the conditions of Turkey and, more importantly, the detection of network speckle disease with ANN and machine learning models has been achieved with high success compared to the literature.

In this study, original development has been made using classification and regression models using climatic parameters for the prediction of net blotch disease in barley. Some climatic parameters (temperature, humidity etc.) that are understood to directly affect the formation of the disease will be used as input and the status of the disease will be estimated. In this way, the occurrence of the disease will be revealed and it will be possible to combat the disease and to implement it quickly. Since the formation of the disease cannot be prevented due to reasons such as global warming, it is essential to combat the disease and for this, it is important to automatically analyze the factors affecting the disease. As a solution to this problem, in this study, machine learning-based classifier design and regression model design using ANN and regression models were carried out using 9 meteorological climate data taken from the areas where the disease is seen in Kırşehir province and the results were tested and presented in this study. Kırşehir province was selected because it is one of the regions where both barley production is high and net blotch disease has been seen intensively in recent years.

The main contributions of this study are that it has been shown that barley net blotch disease can be predicted by analyzing the information obtained from the existing measurements. This information provides important clues that pesticide use can be reduced and productivity can be increased. This prediction was made with a unique artificial neural network design with machine learning models that evaluate nine different meteorological climate data together. The model has been evaluated for the first time in a data-dependent manner in predicting net blotch disease by evaluating data specific to Turkey and different meteorological climate conditions of Turkey. The performance obtained in this study is one of the best results among both studies conducted in Turkey and studies conducted in different countries. In addition, the factors affecting barley net blotch disease were examined by performing tree analysis for the Pearson Correlation Coefficient matrix and XGBoost algorithm, and the effects of each feature on the formation of the disease were evaluated and discussed comparatively.

In the second section of this paper, detailed analysis of the data set used in the study and details of the methodologies used in the testing phase are given. In the third section, both the classification and regression analysis results are presented and discussed comparatively. In the last section, a general summary of the study is made, the results are evaluated and suggestions for future studies are presented.

2. Material and Method

In this study, models were proposed in which nine different meteorological climate data were used as inputs to predict net blotch disease in barley and classification was performed using machine learning and artificial neural network design. The model has the basic steps of collecting data, pre-processing processes, standardizing the data, separating it into training and test samples and classifying it using machine learning and artificial neural network in the final stage and these steps are summarized as shown in Figure 1. Each step is explained in detail in this section.

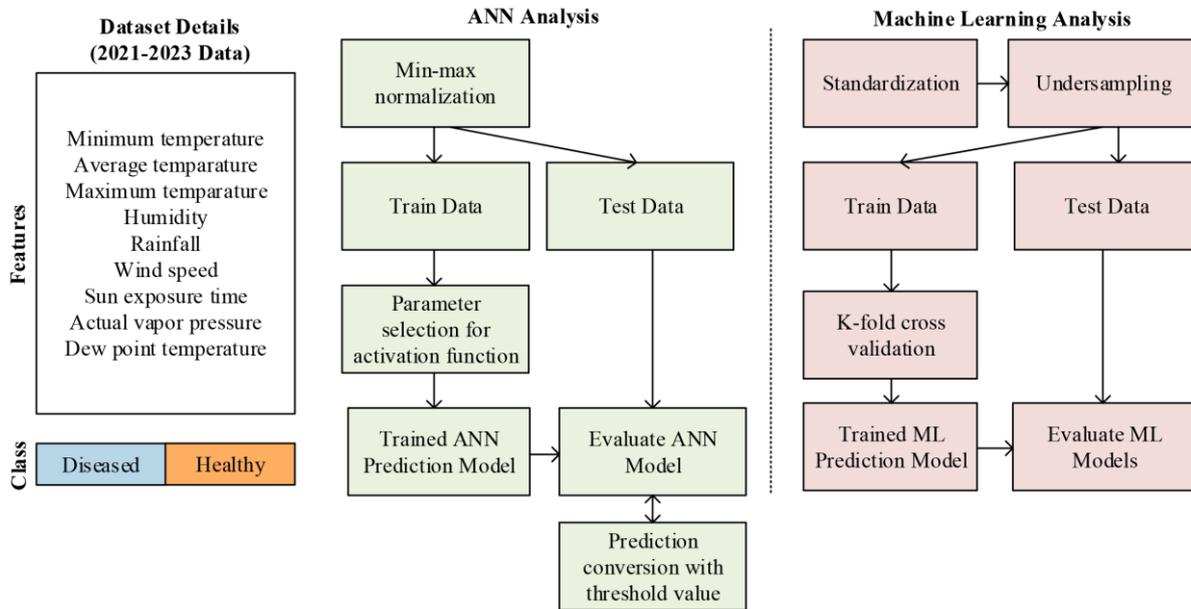


Figure 1. The following steps are undertaken in the process of net blotch prediction

2.1. Dataset details

In this study, a machine learning-based method was proposed to estimate the density of net blotch disease in barley plants in Kırşehir province. The dataset consists of data collected separately for each of 9 meteorological data, 1 data per day, for the years 2021-2023, by taking the average of the data obtained from the two General Directorate of Meteorology stations closest to a selected land in the Kırşehir province in the central Anatolian region of Turkey. The dataset was collected within the scope of this study to use 9 meteorological climate data as input and to estimate the disease as Present (1) or Absent (0) for the relevant province.

The nine meteorological data presented herein were determined through comprehensive literature research. Accordingly, it was stated that minimum, maximum, average temperature and humidity directly affect plant diseases [30], relative humidity facilitates spore formation on the leaf [31], precipitation accelerates disease formation due to creating a moist surface [32], wind speed and direction are important in the spatial spread of the disease [33], and sunshine duration increases disease susceptibility [34]. It is for this reason that the analysis in this study focused on these seven meteorological data sets. Furthermore, two additional data points have been obtained through computational analysis. The calculated climate parameters were obtained from the measured climate parameters [35] using literature sources. The calculated climate parameters are ea (actual vapor pressure), edew (dew point). Each of these variables was selected to increase the predictive power of the model with biological bases and it was observed that they made significant contributions in multivariate analyses. The meteorological station map of the region in which the data was collected is displayed in Figure 2.

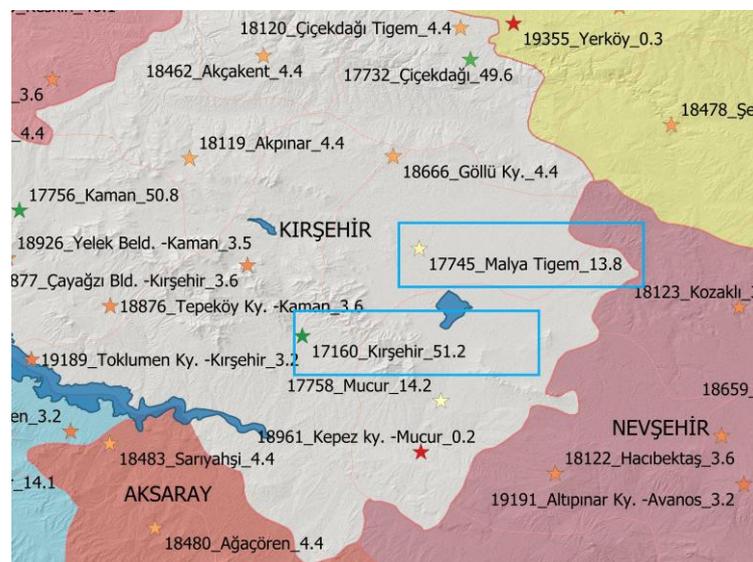


Figure 2. Approximate area of observation areas and stations from which data was collected

The central station of Kırşehir numbered 17160 is located at latitude 39.16393 and longitude 34.1561, and its altitude value is 1007. The Boztepe station of Kırşehir numbered 17745 is located at latitude 39.30385 and longitude 34.34211, and its altitude value is 1127. There are a total of 1095 (365x3) data for a 3-year period taken from these two stations closest to the selected region, and it is based on the daily values of meteorological data taken from Kırşehir province. The variables found in the data collected from these stations and the expressions representing them are given in Table 1.

Table 1. State variables

Variable	Definition	Details
x_1	Minimum temperature	It is the meteorological measurement of the lowest temperature value reached during the day for the relevant date.
x_2	Average temperature	It is the average of temperature values measured at intervals during the day for the relevant date.
x_3	Maximum temperature	It is the meteorological measurement of the highest temperature value reached during the day for the relevant date.
x_4	Humidity	It is the humidity value measured during the day for the relevant date.
x_5	Rainfall	It is the amount of rain, if any, during the day for the relevant date. Since there is no rain for most days, it comes as 0.
x_6	Wind speed	It is the speed value measured during the day for the relevant date.
x_7	Sun exposure time	In case the measured data is incorrect, the equation given below is also used to calculate it and if there is a deviation of more than 10% - 20%, the calculated value is used instead of the measured one. $R_s = k_{RS} \sqrt{(T_{max} - T_{min})} R_a$
x_8	Actual vapor pressure	The actual vapor pressure is calculated from the relative humidity according to the equation given below. $ea = \frac{Rh_{ort}}{100} x e^{oT_{min}}$, Rh_{ort} : Daily average percent relative humidity
x_9	Dew point temperature	The dew point temperature is defined as the temperature at which air must be cooled in order to become saturated. The actual vapour pressure of air is the saturation vapour pressure at the dew point temperature.
y	Disease Status (Diseased or Healthy)	Barley net blotch indicates disease intensity.

Sample values of the collected data are presented in Table 2. As seen in Table 2, there are 9 meteorological data affecting the disease, and the classification results showing the periods in which the disease occurs for each year are considered as the estimated value.

Table 2. Dataset for 2021-2023 years

DataID	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y
1	-3,5	3,6	12,7	68,4	0	0,7	9,96	0,66	1,1	0
2	-1,9	4,8	13,4	61,8	0	1,2	9,84	0,64	0,6	0
3	-1,8	4,2	12,6	66,9	0	1,2	9,32	0,67	1,3	0
4	-2,4	3,4	11,4	73	0	0,8	8,64	0,68	1,5	0
5	-1,4	5,1	12	78,1	1,5	0,6	8,53	0,76	3	0
...
1095	-0,8	3,5	10,8	70,4	0	0,7	9,3	0,66	1,1	0

In order to analyze the relationship between different meteorological data in the dataset, the Pearson Correlation Coefficient matrix shown in Figure 3 was taken. This matrix was prepared to analyze the linear relationships between meteorological variables. The coefficients obtained are obtained by dividing the covariance of two variables by the product of their standard deviations.

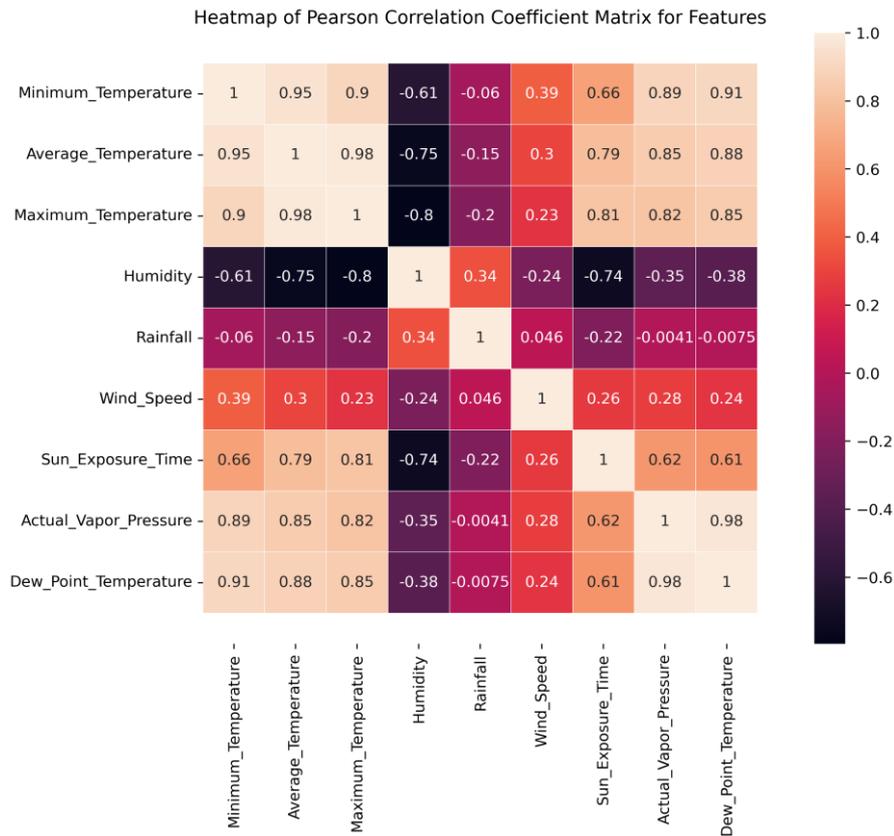


Figure 3. Heatmap of Pearson Correlation Coefficient Matrix for Feature Analysis

When Figure 3 is examined;

- (1) Minimum_Temperature has a very high linear relationship with other temperature variables (Average_Temperature and Maximum_Temperature), indicating that temperature changes are generally affected by common factors. In addition, the strong relationship of low temperatures with high vapor pressure and dew point temperature can be explained by the fact that air reaches saturation faster at low temperatures and dew formation increases. The negative correlation between relative humidity and temperature confirms the fact that relative humidity can be high at low temperatures despite the small amount of water vapor that air can carry.
- (2) Maximum_Temperature has a strong relationship with variables such as sunshine duration, vapor pressure and dew point temperature. This suggests that temperature changes are closely related to water vapor dynamics in the atmosphere and sunshine duration. The strong negative relationship with humidity confirms that relative humidity decreases as temperature increases. However, no significant relationship is observed between precipitation and maximum temperature.
- (3) Average_Temperature has a very strong linear relationship with Maximum_Temperature and both are affected by similar environmental conditions. In addition, variables such as sunshine duration and vapor pressure also have a significant effect on average temperature. The negative relationship with humidity suggests that relative humidity tends to decrease as temperature increases. However, no significant relationship is observed between precipitation and average temperature.
- (4) Humidity has strong negative relationships with sunshine duration and temperature variables, indicating that relative humidity decreases as temperature increases and sunshine duration increases. Weak negative relationships with wind speed and vapor pressure suggest that humidity is also affected by less direct factors. The weak positive relationship between humidity and precipitation suggests that although high humidity creates favorable conditions for precipitation, this effect is limited.
- (5) The low correlation values (positive and negative) of Rainfall make it reasonable to classify it as an independent variable. This shows that Rainfall exhibits a more independent change compared to other variables and is a less susceptible variable in the system under study.
- (6) Wind_Speed stands out as a variable with weak relationships with other variables. Especially the positive correlation of $r=0.39$ with Minimum_Temperature is relatively remarkable, but this relationship is not statistically strong. Therefore, like Rainfall, Wind_Speed can be classified as an independent variable.
- (7) Sun_Exposure_Time shows strong positive relationships with temperature variables, while it has a strong negative relationship with Humidity. It also has moderate positive relationships with Actual_Vapor_Pressure

- and Dew_Point_Temperature. This analysis shows that the duration of sunshine has a significant effect on temperature and increases the amount of water vapor in the air, increasing the vapor pressure and dew point.
- (8) Actual_Vapor_Pressure has strong positive relationships with all temperature variables because the amount of water vapor in the air increases as the temperature increases. These relationships show that the humidity and water vapor in the air change in parallel with the temperature.
 - (9) Dew_Point_Temperature has strong positive relationships with temperature variables (Minimum, Average and Maximum) and an inverse relationship with humidity. It also has a very strong positive relationship with Actual_Vapor_Pressure, since both variables are directly related to the density of water vapor in the air. This analysis clearly demonstrates the close relationship between dew point and vapor pressure and the strong correlation of dew point with temperature.

As a result, temperature (Average, Maximum, Minimum) is generally strongly related to humidity, vapor pressure and dew point. As temperature increases, humidity and vapor pressure decrease, while dew point increases. Sun Exposure Time has a particularly strong relationship with temperature and has an inverse effect on relative humidity. Humidity is generally inversely proportional to temperature and has a particularly negative relationship with sunshine duration. Rainfall is weakly related to vapor pressure and dew point, so its direct effect is minimal. In light of this information, it can be said that factors such as temperature, sun exposure time and humidity have the most significant effect on environmental variables.

2.2. Undersampling

It is very important for machine learning models to have balanced data on a class basis, and when there are classes with unbalanced data, they may show both low performance and convergence to the class with a large number of examples [36, 37]. Undersampling and oversampling methods can be used for data sets with unbalanced samples. In the undersampling method, the process of removing samples from the majority class is applied to balance the number of samples. As a result of this process, the data shows a balanced distribution between classes. At this point, removing samples from the majority classes can cause loss of valuable data. For this reason, if the majority class is well represented, removing some samples will not be a problem.

Within the scope of this study, due to the nature of the problem and the development period of barley, net blotch disease has a feature that is seen in certain months of the year (March-April-May-June) and not in other months of the year. 9 different features are collected annually. For this reason, while there is no disease in an 8-month period due to host-pathogen relationships, this disease is encountered in a 4-month period. However, the start and end times of the disease vary for each year. In this case, the number of samples of the Healthy class reaches approximately twice the number of samples of the Diseased class. In this case, the Accuracy value is high in the tests performed, while the precision and recall values are low, on the contrary. This situation showed that the proposed model converges to a certain class. For this reason, the data was reduced and the clusters were balanced and used in model training. Over-sampling method was not used for data balancing because it can result in over-fitting, where the model learns the noise and variability of the minority class and its poor performance can be seen in new samples. Undersampling was preferred to avoid this.

2.3. Feature scaling

Feature scaling is a critical preprocessing step for machine learning models. There are different methods for feature scaling, such as standardization and normalization. These processes help the model to become more discriminative and efficient, while also helping to identify and manage outliers in the dataset. With this process, variables and attributes are transformed to a similar scale, or the effect of outliers on the overall training process of the model is minimized. Data features are scaled so that they are distributed equally and in a similar range. This process is useful when each feature has different measurement units, ranges, and sizes. If this is not done, the model will converge to some features in a biased manner. This process is necessary for the model to be unbiased.

In this study, 9 different features are used as inputs and these have 9 different ranges of measurement values in 6 different units. The minimum, maximum, average and standard deviation values of each feature are given in Table 3. Each feature is in different units and the value ranges are very different from each other.

Table 3. Details of meteorological data and analysis results

Feature	Unit of measurement	Minimum	Average	Maximum	Standart deviation
x_1	°C	-19.2	7.0	25.6	8.125274
x_2	°C	-14.4	12.7	32.1	9.065925
x_3	°C	-6.6	19.2	39.7	10.04197
x_4	g/m3	17.9	56.7	93.2	15.24041

x_5	mm	0	0.91	46.6	3.127765
x_6	m/sn	0.4	1.76	6.4	0.895951
x_7	Mjoule/m ²	1.37	16.83	29.25	7.327920
x_8	kPa	0.21	0.94	2.04	0.364423
x_9	°C	-13.8	5.11	17.8	6.013500

Due to these different types of features, standardization process was performed in the data within the scope of this study. This technique is a technique where the values are centered on the standard deviation and mean. In this method, the general distribution of the data is not affected and the outliers do not affect the distribution. The general mean of the data is set to zero and the standard deviation is set to 1 [38]. As a result, the data are subjected to the standardization process and used in the training and testing process of the machine learning model.

2.4. Hyper-parameter selection and evaluation metrics

Machine learning models are used today to solve many different problems [39, 40]. In this study, the use of different machine learning techniques for predicting net blotch disease in barley plants and the obtained performance results were analyzed comparatively. The results obtained in these analyses were measured according to the equations given in Equation (1) and shared in detail in the Results section.

$$\begin{aligned}
 Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\
 Precision &= \frac{TP}{TP + FN} \\
 Recall &= \frac{TP}{TP + FN}
 \end{aligned}
 \tag{1}$$

According to these results, hyper-parameter selection process was performed with RandomSearchCV for the XGBoost algorithm which obtained the highest performance. RandomizedSearchCV is based on sampling a fixed number of hyperparameter combinations. Instead of all combinations like GridSearchCV, it evaluates on a more limited sample set. This significantly reduces the cost compared to GridsearchCV. Since the hyperparameter space of our model is large due to the ensemble model structure and contains both continuous and categorical parameters, RandomSearchCV is more suitable. The efficacy of this algorithm is such that it significantly increases the probability of identifying optimal parameter combinations. Furthermore, Bayesian and Genetic algorithms have been shown to be both powerful and efficient in the selection of hyperparameters. However, it should be noted that these algorithms require much more complex structures and longer running times. Consequently, hyperparameter selection was executed via Randomized SearchCV. As a result of these processes, the following values were selected and used: 'subsample': 0.8 [0.6, 0.8, 1.0], 'min_child_weight': 1 [1, 5, 10], 'max_depth': 4 [3, 4, 5], 'gamma': 1 [0.5, 1, 1.5, 2, 5], 'colsample_bytree': 1.0 [0.6, 0.8, 1.0].

3. Comparative Results and Discussion

In this study, in the first stage, machine learning and in the second stage, artificial neural network-based classification were performed and the results were evaluated comparatively. In Table 4, the prediction results obtained for 10 different machine learning models are given by measuring for 4 different metrics. Accordingly, the highest performance was obtained with the XGBoost algorithm with an accuracy value of 88.9%. 9 different input parameters were evaluated together and the disease status was classified as present or absent. Although the results obtained differed according to the classifiers, the performance was generally at an acceptable level. This success was achieved in tests conducted with data collected for a 3-year period, giving clues that higher success can be achieved when data is collected for a longer period.

Table 4. Classification Results of Net Blotch Prediction with Machine Learning

Classifier	Accuracy	Precision	Recall	F-score
KNN	0.856	0.824	0.848	0.836
Random Forest	0.850	0.787	0.894	0.837
Decision Tree	0.739	0.667	0.788	0.722
Extra Tree	0.889	0.836	0.924	0.878
Logistic Regression	0.797	0.754	0.788	0.770

Gaussian Naive Bayes	0.758	0.688	0.803	0.741
Linear Discriminant Analysis	0.804	0.750	0.818	0.783
Ada Boost	0.876	0.805	0.939	0.867
Gradient Boosting	0.863	0.800	0.909	0.851
XGBoost	0.889	0.948	0.850	0.8970

The ROC curve is a visual representation of how well a classification model performs. The area under the curve (AUC) represents the overall performance of the model, and as shown in the first figure of Figure 4, XGBoost (0.9528), Gradient Boosting (0.9493), and Random Forest (0.9508) outperformed other models by reaching the highest AUC values. Models such as Gaussian Naive Bayes (AUC = 0.8600) showed limited success with low AUC values. Cross-validation is a method used to measure the generalization ability of a model by evaluating its performance on different data segments. The validation results obtained for each algorithm are shown in the second figure of Figure 4 as a box plot. The box plot provides information about the distribution of the data. The line inside the box represents the median, the upper and lower edges of the box represent the quartiles, and the points outside the box represent the outliers. XGBoost, Random Forest, and Gradient Boosting stood out in terms of both performance and consistency with their high median AUC scores and narrow distributions. In contrast, models such as Gaussian Naive Bayes produced lower and inconsistent results. This analysis reveals that XGBoost and similar models are more reliable and effective than other methods, especially in classification problems.

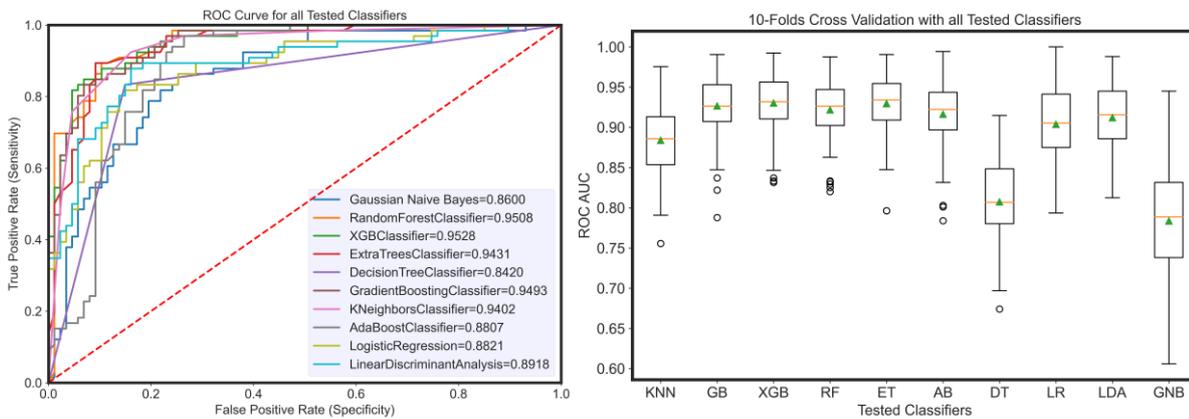


Figure 4. ROC Curve and 10-Folds Cross Validation with all Tested Classifiers

Different analyses were performed to evaluate the results obtained for the XGBoost algorithm, which achieved the highest performance. Accordingly, when the confusion matrix given in Figure 5 is examined, in the tests conducted with a total of 306 samples for each class, the predicted value of the cases where the barley is healthy is high, while the predicted value of the cases where I am sick is lower. For this reason, the TN value is low. The main reason for this situation is that the period when the disease is seen during the year is in a low period compared to all days of the year. This situation makes it difficult to learn the sick periods. It is evaluated that the difference in performance between classes will decrease when the amount of data is collected for longer periods (such as 20-30 years) as in the literature. In addition, since the results are taken between classes with equal sample numbers, it also shows that the model does not converge to any class.

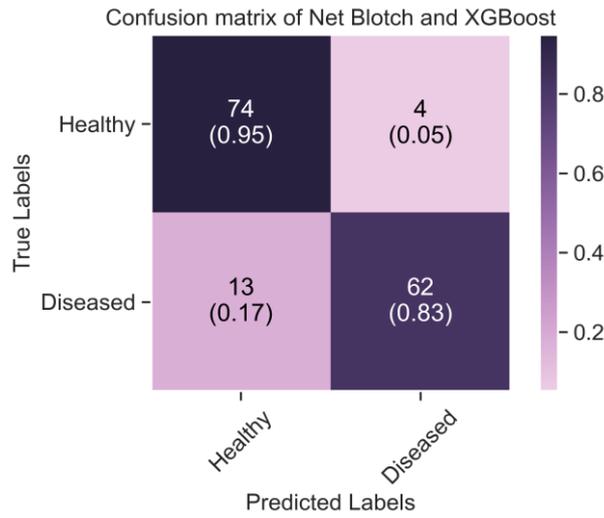


Figure 5. Confusion matrix of XGboost for Net Blotch Disease Detection

Figure 6 shows each AUC value obtained after 10-Fold CV using the XGBoost classifier. The ROC curve represents the balance between sensitivity and specificity. The results show that the classifier can correctly determine the presence of net blotch disease in barley independently of data selection and has high sensitivity. The average AUC value is 0.928, which is quite high compared to the literature. In addition, when the precision-recall curve is examined, it shows that it has high precision and recall values. This proves that the model is low in data dependency and that the expected performance can be achieved in all cases.

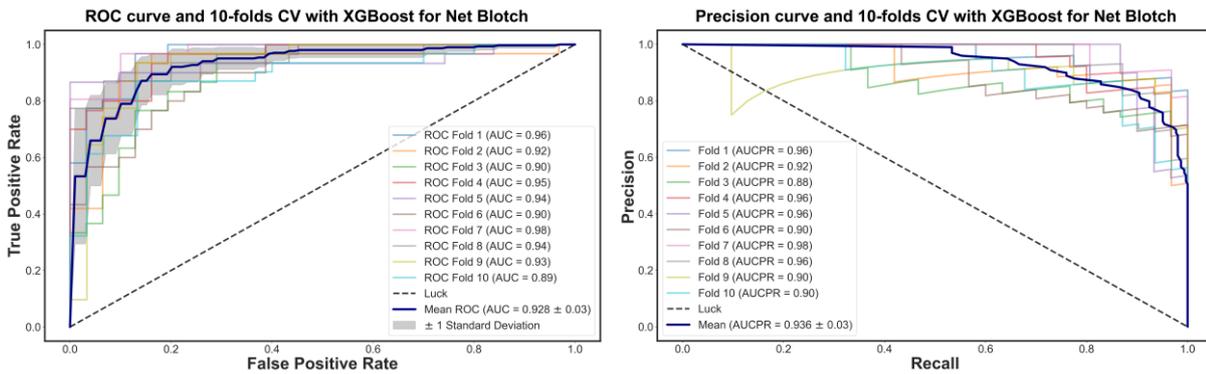


Figure 6. ROC Curve and Precision Curve for 10-folds CV with XGBoost for Net Blotch Disease Detection

The decision tree structure shown in Appendix B for the XGBoost algorithm was structured by selecting Feature 6 (sun exposure time). This selection was made with the aim of minimizing the impurity value (error rate at each node during branching). The aim of the model is to correctly classify the target value by reducing the impurity value to zero at the end of each branch. Feature 6 (sun exposure time), in addition to being selected as the first node, was also used repeatedly in the sub-branches, becoming the main determinant of the decision tree. Branching in the decision tree was made with the features Feature 5 (wind speed) and Feature 0 (minimum temperature). The processes were first started on the left branch with the lowest impurity value, and then the processes continued on the right branch. The model optimized the classification processes by using only those that had an effect on impurity reduction out of a total of 9 features. While Feature 6 (sun exposure time) and Feature 1 (average temperature) were the most used features, Feature 5 (wind speed) and Feature 3 (humidity) also created significant branches. Feature 2 (maximum temperature) and Feature 8 (dew point temperature) were excluded because they did not provide sufficient contribution to the classification. As a result, the decision tree effectively reduced the impurity value and performed correct classification by using only the necessary features to obtain the correct target value.

In the second stage, an originally designed ANN network was used to predict net blotch disease in barley. While the collected 9 meteorological climate data were used as input, the disease period varied from year to year and the disease status was marked as Disease Present (1) or Absent(0). For this reason, the activation function used in the ANN model was considered non-linear, solved and the obtained values were converted to class values to represent the disease intensity and the results were obtained. Then, a threshold value (0.5) was taken into account for the disease intensity and the disease intensity was converted back to class. Thus, it was possible to compare the obtained results with machine learning models and other studies in the literature. In this study, the tests were

carried out comparatively and with the same metrics in order to reach the best model, which is the main purpose and to reveal the best disease prediction model.

The ANN model presented in this study and whose scheme is given in Figure 7 was used to predict the classified Network speckle disease status using nine different input parameters.

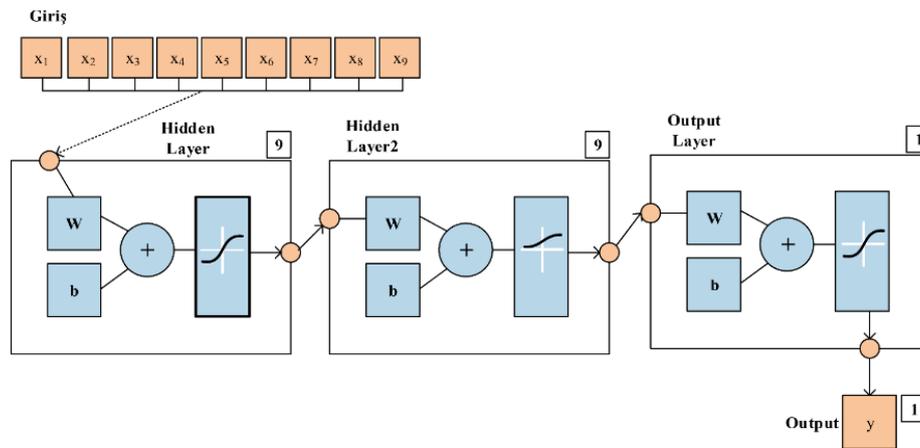


Figure 7. ANN model structure

The selection of activation functions is critical in network design and facilitates the discovery and learning of complex patterns and relationships found in the data by applying them to the output of each neuron. It also ensures that the network maintains its nonlinear structure. The layers of the network, the number of nodes in the layers, and the activation function chosen in each layer directly affect the overall performance of the model [41, 42]. The structure proposed in this study is a non-linear activation function network with tanh, logsig and tanh functions, which has the highest performance as a result of the tests. In addition, the number of dimensions in the hidden layers in this architecture was determined as [9,9,1], respectively, as tested in previous tests.

The y variable was estimated using 9 input variables with the Matlab R2023a program. For the estimation, training and testing processes were performed with the data in Table 5. In addition, 942 rows of the total 1095 rows of data were separated for training and 153 rows for testing. This test data is compatible with the test data of the machine learning model. Thus, the comparison was carried out under fair conditions. Some outputs obtained during the analysis process of the proposed ANN model are given in Table 5.

Table 5. Training Progress and State of ANN analysis.

	Initial Value	Stopped Value	Target Value
Epoch	0	10000	10000
Elapsed Time	-	00:00:17	-
Performance	0.207	0.0325	3e-15
Gradient	3.1	0.0709	0
Mu	0.001	1	1e+10
Validation Checks	0	0	1.11e+4

Some outputs obtained during the analysis process of the proposed ANN model are given in Table 5. Model parameter selections were determined by taking into account the situations where the best results were obtained as a result of optimization. Figure 8 shows the R value graph obtained as a result of training according to these parameters.

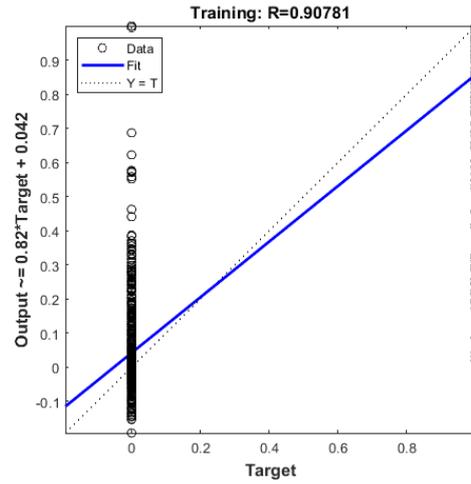


Figure 8. R value for training in ANN analysis

This high R value obtained in the training data shows how well the model has learned the relationship between the target output and the input data. A higher R value shows that different variations in the training data establish a stronger relationship with different variations in the target output. This shows that the proposed ANN model is well adapted to the training data.

In addition, the relevant activation function in the ANN analysis is given in equation (1).

$$y = \text{tansig}(b_3 + LW_2 \text{logsig}(b_2 + LW_1 \text{tansig}(b_1 + IW X))) \quad (1)$$

Here $X(\text{INPUTS}) = (x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6 \ x_7 \ x_8 \ x_9)^T$ and the coefficients are:

$$b_1 = (5,4471 \ 4,805 \ -2,7099 \ 4,7572 \ -1,5805 \ -0,2770 \ 0,4226 \ -1,9653 \ 0,3428)^T,$$

$$b_2 = (3,5345 \ -2,8319 \ 1,5692 \ 0,7199 \ -0,0616 \ 1,1045 \ -2,1216 \ -2,4655 \ -3,6557)^T,$$

$$b_3 = (0,9282),$$

$$IW = \begin{pmatrix} 0,3901 & -0,7792 & -0,0153 & 0,0049 & 0,1003 & -0,5510 & 0,1963 & -1,1922 & -0,0252 \\ -0,4822 & 0,5021 & 0,2103 & -0,0215 & 0,1154 & -0,3759 & 0,2160 & -0,9859 & 0,2997 \\ 0,1159 & 0,5299 & 0,9827 & 0,0833 & -0,1964 & 1,0989 & -0,9330 & -1,0665 & -1,2620 \\ -0,7128 & 0,7959 & 0,2689 & 0,0527 & -0,4098 & -0,8873 & 0,5302 & -0,3205 & 0,7706 \\ -0,8767 & -0,5022 & -0,1114 & -0,2190 & 0,2425 & 0,6175 & 1,0166 & -1,6550 & 1,0473 \\ 1,2060 & 1,7297 & -2,1751 & -0,1534 & -0,3262 & -0,7306 & 0,7672 & -0,5627 & 0,2924 \\ 0,1960 & 2,0286 & -1,3230 & -0,1683 & 0,2963 & 0,8855 & -0,0065 & -0,6400 & -0,5954 \\ 0,3133 & 0,2286 & -0,6683 & -0,0127 & 0,0416 & 0,2422 & 0,6500 & -2,5803 & 0,3516 \\ 0,0440 & -0,3487 & 0,2013 & 0,0185 & -0,0026 & -0,1682 & 0,0228 & -1,4507 & 0,0825 \end{pmatrix},$$

$$LW_1 = \begin{pmatrix} -0,8304 & -0,7507 & 1,7760 & 1,7031 & -1,8692 & 1,4358 & -0,0321 & 1,1556 & 0,1137 \\ 1,3660 & 0,5592 & 1,8655 & -0,3677 & 2,4072 & -3,1495 & 1,3854 & 1,5443 & -1,2991 \\ -0,8381 & -1,8715 & 0,3668 & -2,2207 & 1,9107 & -1,8551 & 1,5325 & -0,0758 & -1,6495 \\ -1,3819 & 0,6448 & 1,4565 & 0,5611 & 2,1756 & 2,0207 & 0,7178 & -0,7040 & -0,3805 \\ 0,5060 & 0,4083 & -1,8738 & 1,8287 & -1,5958 & 0,4125 & 0,6333 & -0,4401 & 1,6109 \\ 0,4869 & -1,4370 & -1,5509 & 0,9669 & 1,6395 & 1,3575 & -0,2195 & 1,2841 & 1,4772 \\ -1,6150 & 2,0476 & -0,2727 & -0,3639 & -0,2249 & 0,2486 & 1,0711 & 0,0543 & -0,7785 \\ -3,4969 & 0,5142 & -0,0606 & 2,4615 & -0,1304 & -0,7784 & -2,4013 & -4,5164 & -1,2127 \\ -1,1593 & 0,9848 & 0,8288 & -0,8891 & 1,5760 & 1,4445 & 2,3380 & 0,6768 & -0,3116 \end{pmatrix}$$

and

$LW_2 = (0,8293 \ 2,3447 \ 2,3641 \ 1,1594 \ 0,3742 \ 0,5540 \ -1,1790 \ -2,4016 \ -2,6664)$ was obtained.

Thus, the prediction results of the proposed model for the network spot disease are given in Appendix-A, where y is the real value, y_p is the value estimated by ANN with the help of the activation function shown in (2) and y_{pc} is the converted values of y_p value to 0 and 1 according to the disease status. According to these values, the confusion

matrix obtained with the proposed ANN model is as given in Figure 9. According to the obtained results, it was understood that the model had a high success rate of 96% in predicting the healthy periods of barley and made an error for only 3 tests. In addition, the ANN model was able to predict the diseased periods with high success according to the literature.

While the accuracy value of the ANN model was 92.16%, the precision, recall and f-score values were 96.10%, 89.16% and 92.50%, respectively. These values are considered successful only for a 3-year data period.

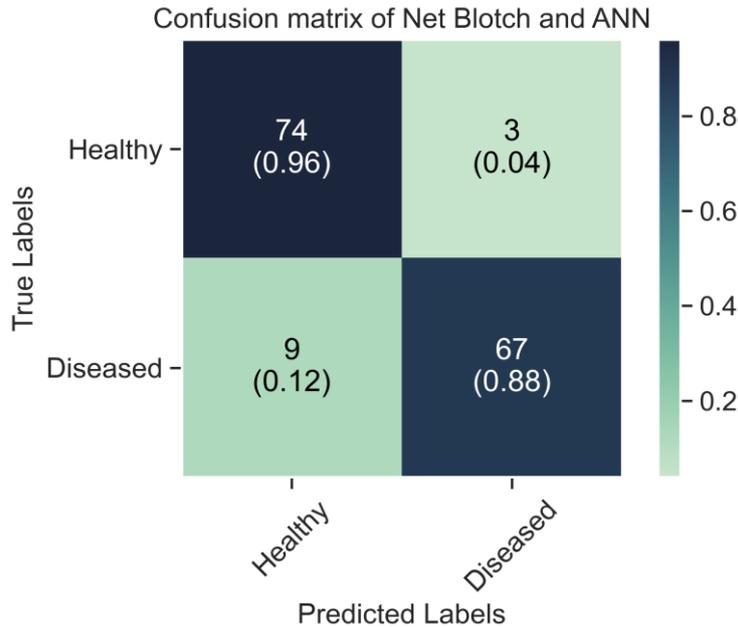


Figure 9. Confusion matrix of ANN for Net Blotch Disease Detection

4. Discussion and Limitations

It is a worldwide known barley leaf disease caused by *P. teres f. teres* (net form, NFNB) and *P. teres f. maculata* (spot form, SFNB). The disease can cause yield losses [43, 44] and grain quality reduction [45]. In this study, different methodologies for the detection of this disease were investigated and evaluated comparatively.

The tests evaluated 9 meteorological climate data collected for a 3-year period between 2021-2023. The presence of the disease for a specific region was analyzed and the region under consideration is a place where net blotch disease has been commonly seen in recent years. The disease starts at certain times (periods) in a cycle every year, shows an increasing trend for a certain period of time and then decreases and ends with the change of climatic conditions. These starting, increasing and ending periods are directly affected by meteorological data. In the tests conducted, while net blotch disease in barley was obtained with the highest success in machine learning models for the relevant region with the XGBoost algorithm, this value is relatively low compared to the ANN model.

According to the analyzed features, the most distinguishing features are those related to the temperature value (minimum, average, maximum temperature etc.). Since the dew point temperature is generally related to the ambient temperature, it is understood that it is another feature that affects the net blotch formation.

One of the most important uncertainties in the performance of classifiers is the uncertainty of the possible distances between the stations where meteorological climate data are taken and the fields where disease measurements are made. Although the data from the station closest to the relevant field was taken within the scope of this study, this did not eliminate the uncertainty. For this reason, it would be useful to apply a method such as the average of the 3 closest stations, etc.

In addition, studies on the transferability of the data-based approach collected in a specific area in Kırşehir to regions with other climatic conditions or to different countries should be continued. It will be possible to increase the widespread impact of the study by making measurements in different climatic conditions where net blotch disease can be seen in barley. The proposed methodology can be used to predict other plant diseases related to the events and the ability to process a database related to the relevant disease will enable the generalization of the proposed methodology.

In this study, machine learning and artificial neural network based models were proposed and used. In addition, it is also possible to use LSTM, RNN, GRU [46] like feedback neural networks and mathematical modeling methods with Difference or Differential Equations, which are widely used in problem solutions. In this way, it will be possible to increase the obtained performance.

The classifiers used in this study have shown that although they contain model complexity to provide the development of plant disease prediction and warning systems, they can learn meteorological features and detect the disease with high success. The obtained results prove that meteorological climate data can be used to detect net blotch disease in barley in Turkey. It is possible to make a clear prediction as whether the disease is present or absent. The proposed model can be integrated with public meteorological services to continue continuous learning and to put it into operation as a warning system for the detection of plant diseases. In addition, as a result of the automatic processing of meteorological data, it is possible to inform and warn farmers and users with text messages or mobile notifications as soon as the exact environmental conditions for the formation of the disease are detected. In this way, producers can optimize the use of chemicals in the fight against the disease, change the planting period of the product or prevent unnecessary use of drugs. In addition, it is possible to work as a decision support system and to assess disease risk in large production areas. It is possible to create regional and country-wide risk maps by integrating the model with satellite data or sensor-based agricultural monitoring systems. Such applications can contribute to the spread of sustainable farming practices and contribute to the reduction of environmental impacts. The model structure proposed utilises meteorological data (temperature, humidity, etc.) as the primary input, eschewing any visual or pictorial elements. All of these data can be obtained from the field and meteorological stations. Consequently, farmers can utilise the model without the necessity for specialised equipment or costly data collection processes, obviating the requirement for any specific hardware. The model exhibits a low computational cost and can be readily executed on online platforms or utilised with laptops. This approach facilitates the dissemination of information to a broad spectrum of users.

In the future, studies can be continued on collecting meteorological data for a longer period and optimizing the model so that the model can be easily applied and used. Model development studies will be carried out first in different geographical regions where barley is cultivated and then on cereal products such as wheat that have similar characteristics to barley. In addition, this study, which includes data from 2021-2024, will be re-tested and its performance will be evaluated with data from 2025 and 2026 to be collected in the coming years. In this way, the durability of the results obtained and their widespread impact on agricultural practices will be confirmed. In addition, the percentage effect of 9 climatic meteorological data on net blotch disease in barley will be analyzed with ANN with normalized data. In this way, the extent to which the input variable affects will be determined numerically.

5. Conclusion

In this study, a comprehensive artificial intelligence analysis was conducted where meteorological climate data and net blotch disease observations were evaluated together and the results were presented comparatively. The research results confirm that meteorological climate data have a significant effect on the occurrence of the disease. 9 different meteorological features were tested with both machine learning and artificial neural network models and the highest recognition success was achieved. The input feature analysis supports the idea that it has a complex effect on the disease and that these features have some dependencies among themselves. The analysis was carried out with data from Kirsehir province in Turkey for a 3-year period. It was considered as a 2-class problem as Disease Present and Absent and the result was a development aimed at detecting the presence or absence of the disease rather than a specific intensity. The aim is to clearly detect the presence of the disease and to take urgent measures. Thus, it will be possible to use less pesticides in the fight against the disease. The results obtained can be used in the reliable identification, modeling and optimization and reduction of pesticide use of meteorological conditions affecting the disease. The dataset was collected originally for this study. The study demonstrated the effectiveness of historical data analysis and its success in disease prediction. This enables the optimization of pesticide use and the reduction of harmful effects on the environment.

In order to generalize the results obtained, different data sets need to be created and tested. In addition, it is important to collect data for different provinces in Turkey where the same disease is seen in different regions and to repeat the tests. In addition, establishing a structure where, in addition to 9 meteorological features, the soil, the plant's own characteristics or the seed characteristics are also evaluated and more features are trained together will provide higher detection success. The results showed that a well-designed and optimized artificial intelligence model can be a part of protection from the effects that may occur as a result of chemical control and reducing environmental effects in the long term and spreading sustainable agricultural practices.

Author Contributions

Conceptualization, R.S.A and B.D; methodology, R.S.A and B.D.; validation, R.S.A and D.Ç; formal analysis, T.D; investigation, N.A, K.A.Ö ; resources, N.A, K.A.Ö; data curation, N.A and K.A.Ö; writing—original draft preparation, T.D; writing—review and editing, T.D, B.D and D.Ç; visualization, R.S.A. and D.Ç All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This study was supported by Scientific and Technological Research Council of Turkey (TUBITAK) under the Grant Number 124F251. The authors thank to TUBITAK for their supports.

References

- [1] Finch, H., Samuel, A.M, Lane, G.F. 2002. Lockhart & Wiseman's crop husbandry. 10th Edition. Cambridge: WoodHead, 3-25.
- [2] Adeboyejo, F. O., Olatunde, S. J., Rustria, G. A., Azotea, A. N. B., Ostonal, J. M., Reyes, J. A., Oyeyinka, S. A. 2023. Asian fermented cereal-based products. *Indigenous Fermented Foods for the Tropics*, Academic Press, 37-56.
- [3] Food and Agriculture Organization of the United Nations. 2024. Cereal Production. <https://ourworldindata.org/grapher/cereal-production> (Accessed Date: 03.12.2024).
- [4] Uçar, Ö. 2020. The Situation of Cereals Cultivation in the World and Turkey. *New Approaches and Applications in Agriculture*, İksad Publishing, 328-344.
- [5] Shahbandeh, M. 2024. Worldwide production of grain in 2024/25, by type (in million metric tons). <https://www.statista.com/statistics/263977/world-grain-production-by-type/> (Accessed Date: 03.12.2024).
- [6] Turkstat. 2023. Crop Production Statistics, 2023. <https://data.tuik.gov.tr/Bulten/Index?p=Crop-Production-Statistics-2023-49535> (Accessed Date: 03.12.2024).
- [7] Foreign Agricultural Services U.S. 2024. Barley Production. <https://fas.usda.gov/data/production/commodity/0430000> (Accessed Date: 03.12.2024).
- [8] Abbass K, Qasim, M. Z., Song, H., Murshed, M., Mahmood, H., Younis, I. 2022. A review of the global climate change impacts, adaptation, and sustainable mitigation measures. *Environmental Science and Pollution Research*, 29, 42539-42559.
- [9] Sadra, S., Mohammadi, G., Mondani, F. 2024. Effects of cover crops and nitrogen fertilizer on greenhouse gas emissions and net global warming potential in a potato cropping system. *Journal of Agriculture and Food Research*, 18, 1-11.
- [10] Duchenne-Moutie, R.A., Neetoo, H. 2021. Climate Change and Emerging Food Safety Issues: A Review. *Journal of Food Protection*, 84(11), 1884-1897.
- [11] Malhi, Y., Franklin, J., Seddon, N. Solan, M. Turner, M.G., Field, C.B., Knowlton, N. 2019. Climate change and ecosystems: threats, opportunities and solutions. *Philosophical Transactions of the Royal Society B*, 375, 1-8.
- [12] Kojiri, T. Hamaguchi, T. Ode, M. 2008. Assessment of global warming impacts on water resources and ecology of a river basin in Japan. *Journal of Hydro-environment Research*, 1(3-4), 164-175.
- [13] Njuki, E. Nava, N.J., Bravo-Ureta, B.E. 2025. Climate and weather impacts on agricultural productivity, *Encyclopedia of Energy, Natural Resource, and Environmental Economics*, 2, 261-268.
- [14] Nazir, N., Bilal, S., Bhat, K., Shah, T., Badri, Z., Bhat, F., Wani, T., Mugal, M., Parveen, S., Dorjey, S. 2017. Effect of Climate Change on Plant Diseases. *International Journal of Current Microbiology and Applied Sciences*, 7(6), 250-256.
- [15] Shah, M.H., Aktar, S.N., Pramanik, K., Pande, C.B., Ali, G.T. 2024. The impact of climate change on plant diseases and food security. *Biocontrol Agents for Improved Agriculture*, Academic Press, 353-384.
- [16] Çelik, E., Karakaya, A. 2015. Eskişehir ili arpa ekim alanlarında görülen fungal yaprak ve başak hastalıklarının görülme sıklıklarının ve yoğunluklarının belirlenmesi. *Bitki Koruma Bülteni*, 55(2), 157-170.
- [17] Tini, F., Covarelli, L., Ricci, G. Balducci, E. Orfei, M., Beccari, G. 2022. Management of *Pyrenophora teres f. teres*, the Causal Agent of Net Form Net Blotch of Barley, in A Two-Year Field Experiment in Central Italy. *Pathogens*, 11(3), 291-294.
- [18] Fenu, G., Mallocci, F. M., 2021. Forecasting Plant and Crop Disease: An Explorative Study on Current Algorithms. *Big Data Cognitive Computing*, 5(1), 2-3.
- [19] Charaya, M. U., Upadhyay, A. Bhati, H. P., Kumar, A. 2021. Plant disease forecasting: Past practices to emerging technologies. *Plant Disease Management Strategies*, 1-30.

- [20] Landschoot, S., Waegeman, W., Audenaert, K., Damme, P. V., Vandepitte, J., Baets, B.D., Haesaert, G. 2013. A field-specific web tool for the prediction of *Fusarium* head blight and deoxynivalenol content in Belgium. *Computers and Electronics in Agriculture*, 93, 140-148.
- [21] Musa, T., Hecker, A., Vogelgsang, S., Forrer, H. R. 2007. Forecasting of *Fusarium* head blight and deoxynivalenol content in winter wheat with FusaProg. *EPP0 Bulletin*, 37(2), 283-289.
- [22] Shah, D. A., Paul, P., Wolf, E.D., Madden, L. V. 2019. Predicting plant disease epidemics from functionally represented weather series. *Philosophical Transactions B*, 374, 1-6.
- [23] Shah, D. A., Wolf, E.D., Paul, P., Madden, L. V. 2019. Functional Data Analysis of Weather Variables Linked to *Fusarium* Head Blight Epidemics in the United States. *Phytopathology*, 109(1), 96-110.
- [24] Shah, D. A., Molineros, J., Paul, P., Willyerd, K., Madden, L. V., Wolf, E.D. 2013. Predicting *Fusarium* Head Blight Epidemics With Weather-Driven Pre- and Post-Anthesis Logistic Regression Models. *Ecology and Epidemiology*, 103(9), 1-14.
- [25] Jorgensen, L.N., Matzen, N., Ficke, A., Andersson, B., Jalli, M., Ronis, A., Nielsen, G. C. 2021. Using risk models for control of leaf blotch diseases in barley minimises fungicide use – experiences from the Nordic and Baltic countries. *Acta Agriculturae Scandinavica, Section B – Soil & Plant Science*, 71(4), 247-260.
- [26] Henriksen, K., Jorgensen, L. N., Nielsen, G. C. 2000. 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.
- [27] Ruusunen, O., Jalli, M., Jauhiainen, L., Ruusunen, M., Leiviska, K. 2020. Advanced Data Analysis as a Tool for Net Blotch Density Estimation in Spring Barley. *Agriculture*, 10(5), 179.
- [28] Ruusunen, O., Jalli, M., Jauhiainen, L., Ruusunen, M., Leiviska, K. 2020. Identification of Optimal Starting Time Instance to Forecast Net Blotch Density in Spring Barley with Meteorological Data in Finland. *Agriculture*, 12(11), 1939.
- [29] Ruusunen, O., Jalli, M., Jauhiainen, L., Ruusunen, M., Leiviska, K. 2024. Linear Discriminant Analysis for Predicting Net Blotch Severity in Spring Barley with Meteorological Data in Finland. *Agriculture*, 14(10), 1779.
- [30] Elan, Y., Pertot, I. 2014. Climate change impacts on plant pathogens and plant diseases. *Journal of Crop Improvement*, 28, 99-139.
- [31] Tho, K.E., McCann, E. B., Wiriyaitsomboon, P., Hausbeck, M. K. 2019. Effects of Temperature, Relative Humidity, and Plant Age on Bacterial Disease of Onion Plants. *Plant Health Progress*, 20, 200-206.
- [32] Bowers, J. H., Sonoda, R. M., Mitchell, D. J. 1990. Path coefficient analysis of the effect of rainfall variables on the epidemiology of *Phytophthora* blight of pepper caused by *Phytophthora capsici*. *Ecology and Epidemiology*, 80(12), 1439-1447.
- [33] Xu, R. 2009. Effects of Prevailing Wind Direction on Spatial Statistics of Plant Disease Epidemics. *Journal of Phytopathology*, 149, 155-166.
- [34] Austin, C. N., Wilcox, W. F. 2012. Effects of sunlight exposure on grapevine powdery mildew development. *Phytopathology*, 102(9), 857-867.
- [35] Allen, R. G., Pereira, L. S., Raes, D., Smith, M. 1998. *Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper*. Rome: FAO - Food and Agriculture Organization of the United Nations, 17-28.
- [36] Arslan, R. S., Ulutaş, H., Köksal, A. S., Bakır, M., Çiftçi, B. 2022. Automated sleep scoring system using multi-channel data and machine learning. *Computers in Biology and Medicine*, 146, 105653.
- [37] Arslan, R. S. 2023. Sleep disorder and apnea events detection framework with high performance using two-tier learning model design. *PeerJ Computer Science*, 9, e1554.
- [38] Köksoy, O., Daşbaşı, B. 2005. Ağır Kuyruklu Dağılımlarda Monte Carlo Simulasyonu ile Konum Parametresinin Analizi. *Journal of Arts and Sciences*, 4, 1-14.
- [39] Uğurlu, M., Doğru, İ. A., Arslan, R. S. 2021. A new classification method for encrypted internet traffic using machine learning. *Turkish Journal of Electrical Engineering and Computer Sciences*, 29(5), 2450-2468.
- [40] Hızlısoy, S., Çolakoğlu, E., Arslan, R. S. 2022. Speech-to-Gender Recognition Based on Machine Learning Algorithms. *International Journal Of Applied Mathematics Electronics and Computers*, 10(4), 84-92.
- [41] Daşbaşı, B., Barak, D., Taşyürek, M., Arslan, R. S. 2024. Using an Artificial Neural Networks Approach to Assess the Links Among Environmental Protection Expenditure, Energy use and Growth in Finland. *Journal of the Knowledge Economy*, 1-29.
- [42] Hızlısoy, S., Arslan, R. S., Çökük, G. Ö., Daşbaşı, B. 2024. ANN Analysis and ODE Model of Selected Factors Affecting Mental Health Problems of Children in Turkey. *European Journal of Science and Technology*, 54, 32-45.
- [43] Jayasena, K., Burgel, A. V., Tanaka, K., Majewski, J., Loughman, R. 2007. Yield reduction in barley in relation to spot-type net blotch. *Australasian Plant Pathology*, 36, 429-433.
- [44] Murray, G. M., Brennan, J. P. 2010. Estimating disease losses to the Australian barley industry. *Australasian Plant Pathology*, 39, 85-96.

- [45] Beckes, A., Guerriero, G., Barka, E. A., Jacquard, C. 2021. *Pyrenophora teres*: Taxonomy, Morphology, Interaction With Barley, and Mode of Control. *Frontier Plant Science*, 12, 1-18.
- [46] Arslan, R. S. 2022. FG-Droid: Grouping based feature size reduction for Android malware detection. *PeerJ Computer Science*, 8, e1043, 1-24.

Appendices

Appendix A. Test data and prediction values of the ANN model

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y	y_p	y_{pc}	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y	y_p	y_{pc}
-0,9	2,5	12,2	75,1	4,6	3,1	12,07	0,75	2,9	1	0,0024	0	7	17,3	23,9	35,5	0	1,9	25,28	0,7	1,9	1	0,9995	1
-6,2	-3,2	3	71,2	0,1	1,4	10,55	0,41	-5,4	1	-0,0165	0	6,1	13,7	21,5	59,7	0,9	1,7	23,8	1,04	7,5	1	0,9998	1
-6,1	2	11,3	55,9	0	0,8	16,32	0,48	-3,3	1	0,9946	1	9,7	12,3	17,4	67,7	3,8	1,3	17,54	1,08	8,1	1	0,9974	1
-3,2	4,8	13,5	61,2	0	1	17,02	0,62	0,2	1	0,9915	1	11,5	18	27,6	52,9	0	1,6	25,69	1,34	11,3	1	0,9994	1
-1,3	4,9	12,7	47	0	4,3	17,22	0,48	-3,3	1	0,9982	1	9,4	14,2	20,8	65,3	0,1	1	21,83	1,19	9,5	1	0,9723	1
1,8	3,7	7,7	72,2	0,6	1,4	9,95	0,63	0,4	1	0,0871	0	9,9	16,1	23,6	59	2,5	1,3	25,09	1,22	9,9	1	0,9921	1
2,3	6	15,5	68	4	3,4	15,47	0,84	4,5	1	0,5153	1	10,4	16,9	23,9	60,4	0	1,3	24,14	1,28	10,6	1	0,6808	1
6,3	8,1	11,2	85,7	2,7	0,8	9,79	0,98	6,7	1	0,8314	1	10,3	16,1	24	70,7	20,9	1	24,48	1,5	13	1	0,9997	1
-0,2	5,8	12,1	69,3	0	1,6	16,06	0,7	1,9	1	0,9629	1	13,8	17	21,6	62,3	0,8	3,3	18,56	1,3	10,8	1	0,2484	0
7,1	8,5	12,1	85,7	46,6	1,4	10,59	1,04	7,5	1	0,9902	1	14,4	16,1	19,7	75,2	0,7	0,8	15,38	1,48	12,8	1	0,864	1
-2	0,9	4,5	79,1	0,2	0,9	12,48	0,54	-1,7	1	0,9146	1	-3,5	3,6	12,7	68,4	0	0,7	9,96	0,66	1,1	0	-0,0102	0
2,4	6,3	11,5	73,9	0,1	3,6	15,24	0,77	3,2	1	0,9985	1	0,8	2,1	4,2	92,7	2	0,8	4,54	0,68	1,5	0	0,0053	0
2,3	10,9	16,3	58,1	0,6	2,2	19,45	0,75	2,9	1	0,9788	1	5,7	8,7	14,3	51,7	1	2,9	8,05	0,66	1,1	0	-0,0043	0
4	13,2	20,7	39	0	1,8	21,54	0,63	0,4	1	0,9592	1	5,1	11,8	18	42,6	0	2,4	13,59	0,63	0,4	0	0,0007	0
-1,2	3,5	7,3	72,5	0	0,6	15,98	0,57	-1	1	0,9897	1	13,1	21,1	28,4	62,1	0	1,5	26,17	1,67	14,7	0	0,5944	1
4,4	13,9	21,6	49,4	0	1,4	23,28	0,84	4,5	1	0,9997	1	16,6	22,9	30,1	41,7	0	1,7	24,34	1,28	10,6	0	0,1304	0
10,6	15,1	21,7	51,6	0	1,6	19,13	1	7	1	0,5438	1	19,2	29,5	38,7	23,7	0	1,1	27,69	1,08	8,1	0	0,001	0
5,8	16,1	23,5	42,8	0	2	24,66	0,82	4,1	1	0,9997	1	17,8	25,7	32,9	35,5	0	3,1	26,87	1,25	10,3	0	0,0026	0
7,2	19	26,9	41	0	1,6	26,51	0,93	5,9	1	0,9997	1	15,2	20,6	28,6	52,2	0	4,3	21,66	1,47	12,7	0	-0,0947	0
10,3	19,9	27,8	48	0	1	25,43	1,2	9,7	1	0,9752	1	17,2	27,7	36,5	26,2	0	1,4	23,63	1,06	7,8	0	-0,4249	0
7,9	13,7	19,9	65,4	0,5	1	21,38	1,11	8,5	1	0,96	1	13,2	22	30,4	34,1	0	1,9	22,16	1	7	0	-0,0824	0
4	10,5	17,9	42,5	0	2,9	27,03	0,61	0	1	1	1	14,4	16,1	21	66,4	0	2,5	11,73	1,37	11,6	0	-0,1417	0
9,1	20,9	30,9	32,8	0	1,3	27,37	0,92	5,8	1	0,9997	1	5,4	11,5	18	53,5	0	1,9	14,36	0,79	3,6	0	-0,0077	0
7,2	20,1	27,7	40,9	0	2,3	27,27	0,97	6,5	1	0,9997	1	5	12,4	19	65,2	0	1,1	13,3	1	7	0	0,017	0
10,2	13,9	19,5	57,2	0	2,4	19,78	1	7	1	0,9336	1	4,1	9,7	18	64,9	0	0,9	11,56	0,93	5,9	0	-0,0415	0
13,3	22,7	30,9	32,7	0	2,5	28,6	0,98	6,7	1	0,5657	1	-4,1	4,1	13,6	66,2	0	1,1	12,01	0,67	1,3	0	-0,0179	0
8,2	18,8	27,3	48,4	0	1,2	27,98	1,14	8,9	1	0,9997	1	5	7,4	12,3	79,5	0	0,8	6,63	0,91	5,6	0	-0,0156	0
7,7	14	21,1	55,8	0	1,6	27,54	0,99	6,8	1	0,9978	1	4,9	5,5	10,2	79,5	1,3	1	5,27	0,84	4,5	0	0,0033	0
6,9	18,5	26,8	42,4	0	1,3	28,28	0,96	6,4	1	0,9997	1	-9,6	-7,1	-1,6	78	2,4	1,1	6,35	0,32	-8,6	0	-0,0062	0
12,7	18,3	24,7	62,6	0,2	1,6	23,13	1,43	12,3	1	0,5997	1	-1,1	3,1	10,8	72	0	0,9	8,02	0,67	1,3	0	-0,0035	0
10,4	17,8	25,6	57,7	0	1,6	25,3	1,31	11	1	0,9965	1	-10,5	-4	1,9	67,9	0,3	1,1	10,71	0,33	-8,2	0	-0,0021	0
-3,6	0,8	6,3	69,6	0	1,3	15,04	0,49	-3	1	0,7773	1	-14,9	-5,1	-1,2	85	0	0,7	10,46	0,32	-8,6	0	-0,1191	0
-3	7,2	16,6	49,1	0	1,1	20,5	0,58	-0,7	1	0,9964	1	-6,5	0,3	8	73	0	1	14,24	0,53	-1,9	0	0,0166	0
5	17	27	29,3	0	1,8	23,86	0,65	0,9	1	0,9983	1	4,2	6	10	81,7	7,2	1,5	8,92	0,84	4,5	0	0,1186	0
5,9	15,7	23,1	32,7	0	2,9	21,71	0,61	0	1	0,7167	1	0,5	0,2	2,5	85,9	17,8	2,4	5,57	0,58	-0,7	0	0,0308	0
1,8	12,4	22,6	49,3	0	1,1	23,36	0,85	4,6	1	0,999	1	-1,3	-0,1	2,9	73,8	0,2	4	8,64	0,48	-3,3	0	0,145	0
-3,3	2,6	8,5	48,8	0,3	1,1	18,94	0,39	-6	1	1	1	15	22,8	29,5	48,5	0	3	28,03	1,41	12,1	0	-0,0736	0
3,6	13,1	20,3	45,6	0	0,8	23,08	0,72	2,3	1	0,9999	1	16,2	20,1	25	62,7	0	3,1	19,76	1,57	13,7	0	0,058	0
3,6	11,5	19	46,5	0	1,8	22,9	0,7	1,9	1	1	1	16,3	21,2	27,8	49,3	0	3,9	28,01	1,38	11,7	0	-0,0197	0
8,7	20	30	26,9	0	1,6	25,76	0,72	2,3	1	0,9806	1	20,5	27,3	33,7	33,5	0	2,1	27,13	1,28	10,6	0	-0,0029	0
11,8	19,3	25,8	40,6	0	1,9	22,45	0,95	6,2	1	0,2352	0	20,9	27	32,9	40,7	0	2,2	26,27	1,52	13,2	0	-0,0021	0
10,5	13,9	20,7	77,4	1,7	1,3	19,49	1,44	12,4	1	0,9634	1	21,7	29,1	35,5	30,2	0	2	21,65	1,27	10,5	0	-0,0197	0
6,8	9,4	12,8	70,9	0,8	1,8	15,18	0,88	5,1	1	0,999	1	21,1	25,9	31,7	46,8	0	2,4	17,08	1,68	14,8	0	-0,1788	0
5,7	14,7	21,3	43,1	0	1,9	24,82	0,74	2,7	1	0,9998	1	17,4	25,2	34,9	31,1	0	1	21,12	1,18	9,4	0	-0,0948	0
11,6	14,3	21,1	76,6	3,8	1,6	19,61	1,48	12,8	1	0,9719	1	11,5	16,4	20,8	53,1	3,5	1,2	19,11	1,01	7,1	0	0,7939	1
9,6	9,3	10,2	84,7	9,4	2,9	4,99	1,03	7,4	1	0,5309	1	13	22	30,2	47,2	0	1,2	18,2	1,37	11,6	0	0,0064	0
10,7	14,6	22,9	65,8	0	1,7	22,68	1,34	11,3	1	0,9405	1	12,4	17,2	23,1	50,2	0	1,9	16,6	1,07	8	0	-0,1308	0

15,6	20,3	26,7	49,9	0	2	26,62	1,31	11	1	0,3546	0	12,6	16,5	22,2	66,5	0	1,4	10,9	1,38	11,7	0	-0,0117	0
14,1	23,7	32,3	37,8	0	1,7	27,64	1,22	9,9	1	0,1551	0	6,4	13,7	22,9	42,3	0	1	14,3	0,79	3,6	0	0,0226	0
15,6	24,4	32,3	41	0	1,9	28,83	1,36	11,5	1	0,0086	0	9,9	12,1	13,9	69,5	0	2,2	5,72	0,98	6,7	0	0,0219	0
12,6	20,4	28,3	66,3	12,4	1,1	27,7	1,76	15,5	1	0,9898	1	-1	5,3	12,5	59,7	0	1,3	11,99	0,6	-0,2	0	-0,005	0
13,9	19,5	27,1	60,9	0,7	1,6	24,28	1,58	13,8	1	0,6104	1	7,6	12,8	18,3	49,9	0	1,4	8,33	0,78	3,4	0	-0,0026	0
4,3	10,1	15,3	48,8	0	1,6	13,98	0,63	0,4	1	0,0295	0	1,7	6,2	10,8	69,4	0	1,6	7,04	0,69	1,7	0	-0,0159	0
-0,2	3,8	8,1	76,3	13,4	2,3	12,62	0,64	0,6	1	0,9822	1	5,1	8,7	12,9	68,9	0	0,7	6,34	0,82	4,1	0	-0,0314	0
2	6,3	11,7	74,8	1,6	2,2	14,14	0,78	3,4	1	0,991	1	2,3	6	15,4	78,7	0	0,8	8,14	0,97	6,5	0	-0,0225	0
0,8	6,3	11,5	67,4	0	0,7	15,38	0,68	1,5	1	0,9781	1	-2,2	2	9,4	67,1	0	1	9,95	0,57	-1	0	-0,0019	0
0,1	9,7	17,5	53,7	0	1	20,2	0,7	1,9	1	0,9996	1	-5,6	-2,4	3,3	77,9	0	0,7	6,95	0,46	-3,8	0	-0,0032	0
-5,8	-2,3	2	65,2	3,6	1,4	13,99	0,36	-7,1	1	0,6795	1	0,5	1,9	4,7	68,1	0	2,2	4,97	0,5	-2,7	0	0,0355	0
7,7	13,1	20	53,3	1,1	1,3	18,11	0,9	5,4	1	0,9386	1	-1,6	1,9	6,7	63,9	0	2	7,77	0,49	-3	0	0,0128	0
4	5,2	10,9	76,4	1,1	1,2	13,95	0,81	3,9	1	0,9906	1	-0,2	0,1	1,6	77,7	6,6	1,6	3,83	0,5	-2,7	0	0,0204	0
7,2	8,8	14,3	70,5	10,2	1,1	14,51	0,93	5,9	1	0,9934	1	-5,5	-4,1	-1,3	63,8	3,2	2,8	6,28	0,31	-9	0	0,0524	0
4,6	13	20,4	55,7	0	0,9	23,89	0,91	5,6	1	1	1	-1	3,5	11,6	60,9	0	1,2	12,48	0,59	-0,5	0	-0,0004	0
7,6	14,4	22,5	51,2	0,8	2,1	21,76	0,97	6,5	1	0,9408	1	0,5	7,9	14,9	53,1	0	1,9	15,42	0,62	0,2	0	-0,0225	0
7,6	10,5	16,4	64,7	1,1	2,3	17,31	0,94	6,1	1	0,9987	1	4,7	9,2	13,6	62,5	0,3	1,9	11,99	0,75	2,9	0	0,0522	0
8,5	9,6	13,7	71,9	0,7	1,2	13,56	0,96	6,4	1	0,9987	1	9,3	20,4	27,4	42,1	0	2,3	29,25	1,01	7,1	0	0,035	0
-0,1	9,1	16,4	49	0	1,7	25,16	0,61	0	1	0,9999	1	13,9	22,2	29,6	46,9	0	1,6	27,73	1,35	11,4	0	0,0911	0
10,3	21,5	30	29	0	1,4	19,53	0,8	3,8	0	0,5973	1	18,5	25,5	31,7	44,3	0	2,4	27,3	1,51	13,1	0	0,0964	0
4,9	14,3	22,3	40,8	0	1,5	17,41	0,73	2,5	0	-0,0109	0	16,7	23,6	31,3	31,7	0	1,9	27,66	1,03	7,4	0	0,0937	0
5,9	11,3	19,3	53,8	0	1,3	16,04	0,85	4,6	0	-0,0398	0	13,5	23,2	32,3	39,2	0	1,2	27,22	1,25	10,3	0	0,2907	0
8,1	10,7	16,7	73,7	8,3	0,8	10,31	1,1	8,4	0	0,0046	0	21,2	26,5	33,5	32	0	1,7	25,26	1,23	10	0	0,0054	0
6,4	14,6	21,7	52,5	0	0,8	12,02	0,93	5,9	0	-0,0019	0	25,6	31,8	39,1	17,9	0	2,8	24,65	0,92	5,8	0	-0,0106	0
9	13,2	20,3	58,3	0	1,3	11,94	1,03	7,4	0	0,0179	0	20,3	26,9	33,9	28,5	0	2,3	25,12	1,09	8,2	0	-0,0015	0
10,2	9,5	13	73	3,6	1,3	4,52	1	7	0	0,0368	0	16,1	22,5	29,7	45,6	0	2,9	21,77	1,37	11,6	0	0,0495	0
-0,5	0	2	65,5	4,9	1,5	3,85	0,43	-4,7	0	0,0315	0	12,6	16,8	23,4	41,2	0	4	20,42	0,89	5,3	0	0,2021	0
1,8	7,5	15,9	69,6	0	0,6	10,2	0,87	5	0	-0,0546	0	7,6	15,8	24,6	36,7	0	1,3	21,31	0,76	3	0	-0,1291	0
2	4,5	9,4	74,9	4,1	1	6,21	0,71	2,1	0	0,0047	0	-0,8	3,5	10,8	70,4	0	0,7	9,3	0,66	1,1	0	-0,0025	0
1,5	5	8,5	61,9	5,2	2,3	9,61	0,55	-1,4	0	0,1021	0												

Appendix B. Super Tree for XGBoost (Depth = 3) (shows all depths tree is also loaded as tree.html)

