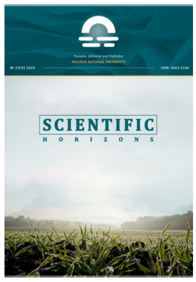
SCIENTIFIC HORIZONS

Journal homepage: https://sciencehorizon.com.ua

Scientific Horizons, 26(2), 31-42



UDC 631.6

DOI: 10.48077/scihor.26(2).2023.31-42

Tomato leaf color as predictor of soil moisture value using machine learning techniques

Svetoslav Atanasov*

PhD Student, Magister in Computer System and Technologies

Trakia University

6015, Students campus, Stara Zagora, Bulgaria

https://orcid.org/0000-0002-2658-1611

Bilyana Harizanova-Petrova

PhD, Chief Assistant Professor

Agricultural University

4000, 12 Mendeleev Blvd., Plovdiv, Bulgaria

https://orcid.org/0000-0001-8437-7718

Radost Petrova

PhD, Associate Professor

Agricultural University

4000, 12 Mendeleev Blvd., Plovdiv, Bulgaria

https://orcid.org/0000-0002-4476-7049

Article's History:

Received: 21.12.2022

Revised: 21.02.2023

Accepted: 9.03.2023

Suggested Citation:

Atanasov, S., Harizanova-Petrova, B., & Petrova, R. (2023). Tomato leaf colour as predictor of soil moisture level using machine learning techniques. *Scientific Horizons*, 26(2), 31-42.

Abstract. Fresh water supplies for irrigation purposes must be used sparingly and judiciously, as water is an invaluable natural resource that is in short supply in much of the Earth. Soil moisture in fields is not uniform everywhere, and deploying thousands of sensors is unnecessarily expensive. The purpose of this publication is to model and predict the relationship between tomato plants leaf color, soil moisture, and thus manage the irrigation process in an optimal manner. The research was conducted using generally accepted methods, the field method, and the method of statistical evaluation of results. Machine learning algorithms (MLA) and data mining are utilized in this paper to model the relationship between RGB color values from tomato leaves and soil moisture and temperature. The color of the leaves of open field tomato plantations grown without stakes is the focus of this study. Three main tasks are fulfilled: to prove that there is a relationship between leaf color and soil moisture, to study its supposedly nonlinear type and to model this relationship with MLA. First, a classifier is trained, and then a model is created and saved. Finally, the efficiency of the chosen model is tested using a different test data set. The name "12-9-6-3" for the methodology of measurements is fgiven. It is proven that the young leaves are more informative about the need for watering. As a result, there is less than a 1% error in predicting soil moisture using the color of tomato leaves

*Corresponding author

considering also soil temperature, using M5P regression model. This predictive model can be used in creation of automated systems for optimal irrigation management and water saving

Keywords: plant water status; soil moisture; leaves color; canopy color; nonlinear estimation

INTRODUCTION

Various methods and sensors for soil moisture measuring are available – direct and indirect, wireless or not. In general, they are expensive and the process of measuring soil moisture is labour-intensive, even overwhelming on the agricultural fields with a huge size of several thousand hectares, for example, or if it is a matter of deploying thousands of sensors. This also includes the non-uniformity of irrigation throughout the entire fields. That is why the relevance of the study is determined by the search for alternative ways to measure soil moisture.

For centuries, farmers have known the condition of their crops by their appearance and colour alone. Experienced farmers and professors from agricultural universities and faculties share their observations based on years of experience that "When a plant (whatever it is – tomato, pepper, wheat, etc.) is dried its leaves turn dark green and when the soil moisture is too high the leaves turn light green" (United Nations Development Programme, 2021).

Irrigation is critical to global agricultural productivity. However, the most effective use of water for irrigation and water-saving efficacy are frequently unexpectedly low (Jägermeyr *et al.*, 2015). It is also estimated that agriculture uses over 70% of global fresh water supplies, with around half of this being lost and wasted (United Nations Development Programme, 2021).

Tomatoes require hundreds of litres of water per square metre. Nederhoff & Stanghellini (2010) note that an average of 300-400 litres of water are used to produce 1 kg of fresh tomatoes in the open field, indicating that the use of irrigation water is extremely high. The authors point out that two other things are specific to watering tomatoes: they require a moderate amount of moisture in the soil. If the watering is overdone, in a short time they form a powerful vegetative mass and less fruit. Another important fact arising from drowning is that tomatoes are quite sensitive to soil deficiency of oxygen.

Koumanov *et al.* (2018) report on two well-known conventional methods of monitoring the irrigation process: groundwater monitoring and plant water monitoring. A number of authors emphasise the following methods of measuring soil moisture: gravimetric approach, dielectric methods (TDR, FDR), neutron probe, tensiometers (Bianchi *et al.*, 2017), thermal probe, soil psychrometer and resistance blocks (Pardossi *et al.*, 2000) are direct and indirect methods of monitoring soil water status (Dobriyal *et al.*, 2012). Some techniques for measuring plant water status include ZIM probes, pressure chambers (Levin, 2019), dendrometers, infrared thermometry measurements (Kalaidjieva *et al.*, 2015), pulse methods (Sela *et al.*, 20007), infrared gas analysers, and porometers.

Some researchers like Afzal *et al.* (2017) utilise electrical capacitance and leaf thickness as indicators of a plant's water condition in predicting plant water status and conducting appropriate watering. Mahan *et al.* (2015) consider canopy temperature as an indicator of plant water status. Ko-Madden *et al.* (2017) employs a proximal leaf monitoring device for determining plant water status to provide precise irrigation in wine grapes. The "sap flow method" measures flow directly through the plant, but the sensor is affixed to the stem and can limit development (Afonso *et al.*, 2020).

The authors of this study suggest that plants themselves can be a kind of biosensor or "high-tech biogadget", an indicator of soil moisture, and therefore, the purpose of the study is to develop an intelligent method for obtaining soil moisture values based on the colour of tomato plants' leaves in the open field.

MATERIALS AND METHODS

The impact of field microclimate variables soil moisture and soil temperature on the leaf colour (Fig. 1) in field tomato plantations (Solanum Lycopersicum, "Nikolina" variety, growing without stake) is the focus of this research. The studies were conducted on the experimental field of Agricultural University of Plovdiv's Faculty of Viticulture and Horticulture (Bulgaria), where all environmental parameters and plants health are kept at their peak. Tomatoes are sown in May. Colour measurements are taken at the start of July, 24 hours before watering. The climatic parameters of the field at that time: the air temperature during the experiments: about 31°C, humidity: 34%, illumination: 75.6 klx. The days preceding the experiment were with extremely high temperatures (about 40°C).

PCE RGB-1002 spectrum colorimeter and environmental parameter measuring equipment PCE-EM 883 was utilised. RGB (red-green-blue) colour space is a basic additive colour model which is mixed with white. Three values describe one colour: red, green, and blue. Each colour percentage can be anywhere between 0% and 100%. The RGB colour space is created using colour perception theories and research.

The Frequency-domain sensors (FDR) for measuring the volumetric water content Onset W-SMC and its pre-calibration to the specific field soil are described by Atanasov (2015). The volumetric water content (VWC) is the ratio of water volume to soil unit volume. The sensor for measuring the soil temperature is Onset WTMB. The field's soil is an alluvial meadow with moisture on FC in the layer 0-40 cm equal to 30.9% VWC (Harizanova-Petrova & Ovcharova, 2015).

Figure 1. *Example of the plants under study*

Source: *photo by the authors*

Based on the authors plant measuring experience, a methodology for conducting the experiment which was used is proposed (Atanasov *et al.*, 2016; Atanasov, 2021), but in this study, the name "**12-9-6-3**" is given: 12 plants studied, 9 plants for model creation (training data set), 6 RGB values from young leaves (from the top of every plant, it's the most upper part), 3 RGB values from old leaves (from the bottom part) for each plant were measured and 3 plants for model testing (testing data set).

The procedure of training classifiers using the training data set consists of three steps. Step one: training a classifier – with a full training data set (9 plants, 54 young and 27 old leaves values). Step two: creating and storing a suitable model and stet three: testing the stored model with additional testing data set (3 plants, 18 young and 9 old leaves values).

Modelling does not account for instances of ill or diseased plants with discoloured leaves. The both testing and training data sets are numerical that trying to predict numerical values, which includes a regression problem. Classifier is trained using entire training data (referred into tables as "all leaves"), as well small datasets filtered using specific criteria (referred into the tables as "young leaves" and "old leaves").

It has been proved that the plant's young leaves are the most informative about the necessity of irrigation (Atanasov *et al.*, 2016; Atanasov, 2021). As a result, twice as many RGB colour samples are collected, and twice as many sample values are displayed in Table 1. At each measurement, the colour values are measured manually and randomly, one by one, from different leaves.

Table 1. *Examples of training data*

Plant No.	Leaf	R	G	B	Soil Mois	Soil Temp	Plant No.	Leaf	R	G	B	Soil Mois	Soil Temp
1	young	111	123	89	17.41	21.72	**5**	old	150	170	105	20.39	23.67
1	young	122	140	100	17.41	21.72	**6**	young	122	142	88	17.74	23.09
1	old	153	176	112	17.41	21.72	**6**	young	135	168	98	17.74	23.09
2	young	128	146	103	15.94	22.68	**6**	old	144	163	107	17.74	23.09
2	young	133	150	102	15.94	22.68	**7**	young	139	159	100	20.44	23.16
2	old	142	161	98	15.94	22.68	**7**	young	130	152	99	20.44	23.16
3	young	106	120	83	14.55	23.85	**7**	old	178	194	108	20.44	23.16
3	young	128	143	87	14.55	23.85	**8**	young	152	167	118	25.56	24.92
3	old	140	156	98	14.55	23.85	**8**	young	143	160	110	25.56	24.92
4	young	108	124	92	18.73	23.50	**8**	old	131	148	99	25.56	24.92
4	young	120	138	92	18.73	23.50	**9**	young	140	158	102	17.29	24.22
4	old	132	151	100	18.73	23.50	**9**	young	143	164	103	17.29	24.22
5	young	130	149	102	20.39	23.67	**9**	old	162	186	122	17.29	24.22
5	young	140	158	114	20.39	23.67							

Source: *developed by the authors*

Table 1 contains only sample input data. The colour measurements are made directly on each random leave with above mentioned pre-calibrated RGB colorimeter, not from photos. Soil moisture and soil temperature are measured once for each plant, near the stem, on 25-30 cm soil depth. Temperature of the soil is in degrees Celsius, and moisture of the soil is in percentages of VWC.

In this study, gravimetric soil moisture was also measured. The equipment is shown in Figure 2. The value obtained for the gravimetric soil moisture in the layer 0-20 cm (the root zone for tomatoes) is 22%. This corresponds to the average value for the obtained volumetric humidity, which ranged from 15% to 22% VWC in the measurements.

***Figure 2**. Equipment for gravimetric soil moisture measuring*

***Source:** photo by the authors*

The models of the relationship soil moisture – leaf colour were created using capabilities of the software applications Statsoft Statistica and Weka (Eibe *et al.*, 2016; Hall *et al.*, 2009). Linear and nonlinear MLAs were used in the modelling.

RESULTS AND DISCUSSION

Normal distribution check. The RGB colour model's Gaussian distribution is verified, and it can be found in the majority of colour components. Conclusions in Table 2 regarding the Gaussian distribution are made using the Shapiro-Wilk test and its significance level (probability). The values for W and p>>0.05 imply that the data is normally distributed. The results reveal that, with the exception of the R component in old leaves, all components have a Gaussian distribution. Table 2 also includes parameters: the arithmetic mean value, the standard deviation, median. The last two parameters: asymmetry and excess show how the distribution is deformed.

***Table 2**. Normal distribution check of RGB colour values; before irrigation*

				Young leaves				
Colour	**Mean**	**SD**	**Median**	**Skew.**	**Kurt.**	**Shap.-Wilk W**	**p**	**Gauss. Distr.**
R	131.17	15.97	132.50	-0.31	-0.83	0.963	0.097	Yes
G	149.63	18.41	151.00	-0.10	-0.41	0.972	0.226	Yes
B	100.46	9.66	102.00	-0.47	-0.31	0.967	0.140	Yes
				Old leaves				
Colour	**Mean**	**SD**	**Median**	**Skew.**	**Kurt.**	**Shap.-Wilk W**	**p**	**Gauss. Distr.**
R	145.78	17.38	142.00	1.02	0.75	0.921	0.042	No
G	164.85	18.18	161.00	0.94	0.89	0.939	0.114	Yes
B	105.93	9.43	107.00	-0.04	-0.36	0.973	0.701	Yes

***Source:** developed by the authors*

Training of classifier and storing of model. Seven regression algorithms for testing have been chosen. When evaluating regression algorithms, two parameters are particularly important: correlation coefficient and the root mean squared error. The correlation coefficient indicates how well the predictions correlate or change in relation to the actual output value. A value of 0 reflects the worst-case scenario, while a value of 1 represents a group of predictions that are completely correlated.

Root-mean-squared error – a metric that indicates how close the model's predictions are to the actual target values. It is the square root of the mean square error, which is the sum of the squared deviations between the predicted and actual values. This statistic can be used to calculate the average error rate of a given prediction. (Brownlee, 2016). In the tables below, three further indications are presented for each MLA: mean absolute error, mean value, and relative absolute error.

Mean absolute error – a metric determining how accurate forecasts or predictions are in terms of actual outcomes. The Relative absolute error (also known as the Root relative squared error) is calculated by dividing the Mean absolute error by the ZeroR classifier error (a classifier that ignores all predictions and chooses the most often occurring value). Amount by which the result differs from the true value is known as absolute error. The relative error is a measure in % compared to the real value.

Applying Zero Rule Regression (ZeroR). This is a fundamental method in the regression algorithm that predicts the mean of the training dataset (Eibe *et al.*, 2016; Brownlee, 2016; Bouckaert *et al.*, 2009). Table 3 shows the obtained results.

Table 3*. Zero Rule Regression Results*

	All leaves	Young leaves	Old leaves
Correlation coefficient	-0.30	-0.33	-0.60
Mean absolute error	2.33	2.36	2.46
Root-mean-squared error	3.06	3.06	3.16
Relative absolute error	100%	100%	100%
Root-relative squared error	100%	100%	100%
The mean value	18.67	18.67	18.67

Source: *developed by the authors*

For the whole set of training data, the ZeroR algorithm predicts a mean soil moisture value of 18.67 (in% volumetric water content) and a root-relative squared error of 3.06. Every other suitable and working MLA must have a greater value.

Linear Regression. When there are more instances than attributes, this technique works well (Witten, 2013). The linear algorithm assumes that the predicted attribute has a linear relationship with the other input attributes. Table 4 shows achieved outcomes.

Table 4*. Linear Regression Results*

	All leaves	Young leaves	Old leaves
Correlation coefficient	0.52	0.51	0.25
Mean absolute error	2.16	2.16	2.43
Root mean squared error	2.61	2.62	3.12
Relative absolute error	92.80%	91.55%	99.02%
Root relative squared error	85.30%	85.72%	98.73%

Source: *developed by the authors*

Thus, the analysed parameters have a confirmed extremely non-linear connection. As a result, authors continue with nonlinear algorithms modelling. **Locally Weighted Learning (LWL) Regression.** Authors use an instance-based algorithm to assign instance weights which are then used by a specified weighted instances handler (Huan *et al.*, 2021). Table 5 shows the outcomes achieved.

Table 5*. Locally Weighted Learning Regression Results*

	All leaves	Young leaves	Old leaves
Correlation coefficient	0.80	0.80	0.79
Mean absolute error	1.39	1.40	1.50
Root-mean-squared error	1.81	1.82	1.84
Relative absolute error	59.54%	59.29%	60.65%
Root-relative squared error	59.31%	59.50%	58.46%

Source: *developed by the authors*

Utilising Model Tree Regression (M5P). The model is a tree, with each leaf containing a linear regression model (Witten, 2013). R. Quinlan created the original M5 algorithm, and Yong Wang improved it. Table 6 displays the acquired results.

There are six linear models (LM) in the nonlinear model:

***Table 6**. Model Tree Regression Results*

	All leaves	Young leaves	Old leaves
Correlation coefficient	0.91	0.84	0.51
Mean absolute error	1.24	1.49	2.06
Root-mean-squared error	1.47	1.80	2.61
Relative absolute error	53.35%	62.90%	83.83%
Root-relative squared error	48.20%	58.67%	82.84%
Number of Rules (Leafs)	6	6	8

***Source:** developed by the authors*

LM num: 1
Soil_Mois=0.0921
*Soil_Temp+14.6772
LM num: 2
Soil_Mois=0.6291
*Soil_Temp+3.0529

LM num: 3
Soil_Mois=1.1355
*Soil_Temp-7.3392
LM num: 4
Soil_Mois=3.8151
*Soil_Temp-70.7297

LM num: 5
Soil_Mois=5.5662
*Soil_Temp-115.8946
LM num: 6
Soil_Mois=4.4335
*Soil_Temp-86.8402

Figure 3 shows the resultant tree depicted graphically after testing with the M5P method:

Decision Tree Regression (REPTree). Decision tree learning is a type of predictive modelling that employs a decision tree (as a predictive model) for progressing from observations about an item (represented by branches) to conclusions about the goal value of the item (represented in the leaves) (Brownlee, 2016). Table 7 shows the collected results.

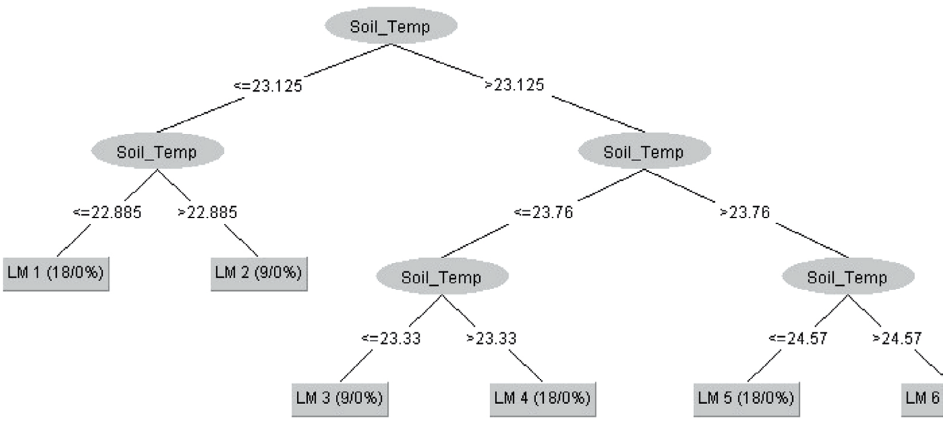


***Figure 3**. Visualisation of the resulting tree after M5P algorithm testing*
***Source:** developed by the authors*

***Table 7**. Decision Tree Regression Results*

	All leaves	Young leaves	Old leaves
Correlation coefficient	0.98	0.83	0.45
Mean absolute error	0.24	0.68	1.85
Root-mean-squared error	0.63	1.69	2.75
Relative absolute error	10.31%	28.84%	75.35%
Root-relative squared error	20.53%	55.11%	87.01%
Size of the tree (rules)	17	7	13

***Source:** developed by the authors*

The tree is turned upside down, beginning at the top or root and progressing through the leaves until a prediction can be made (Brownlee, 2016).

Applying Support Vector Regression (SMOreg). Support Vector Machines (SVM) finds a line of best fit that minimises a cost function's error. These instances are known as support. SVM is a classification method that requires teacher training (Kocaoglu & Akdogan, 2020). Table 8 shows the results obtained.

Table 8*. Support Vector Regression Results*

	All leaves	Young leaves	Old leaves
Correlation coefficient	0.42	0.46	0.44
Mean absolute error	2.18	2.12	2.05
Root-mean-squared error	2.76	2.71	2.85
Relative absolute error	93.69 %	89.61 %	83.55 %
Root-relative squared error	90.32 %	88.66 %	90.30 %

Source: *developed by the authors*

SMOreg is a regression support vector machine implementation (Smola & Schölkopf, 2004). Several algorithms can be used to learn the parameters. The algorithm is chosen by configuring the RegOptimizer.

Multilayer Perceptron Regression. Other names for it include artificial neural networks and neural networks. Neural networks are a tough method to utilise for predictive modelling since there are so many configuration parameters that can only be adequately tweaked by intuition and a lot of trial and error. (Brownlee, 2016). Table 9 shows the outcomes.

Summary table. Table 10 summarises the above calculations in a concise manner

Table 9*. Multilayer Perceptron Regression Results*

	All leaves	Young leaves	Old leaves
Correlation coefficient	0.64	0.48	0.17
Mean absolute error	2.04	2.17	3.29
Root-mean-squared error	2.54	2.83	4.66
Relative absolute error	87.42%	91.95%	133.8%
Root-relative squared error	87.03%	92.48%	147.7%

Source: *developed by the authors*

Table 10*. Summary table*

All leaves	ZeroR	Linear Regr.	LWL	M5P	REPTree	SMOreg	MultyPerc
Correlation coefficient	-0.30	0.52	0.80	0.91	0.98	0.42	0.64
Mean absolute error	2.33	2.16	1.39	1.24	0.24	2.18	2.04
Root-mean-squared error	3.06	2.61	1.81	1.47	0.63	2.76	2.54
Relative absolute error	100 %	92.80 %	59.54 %	53.35 %	10.31 %	93.69 %	87.42 %
Root-relative squared error	100 %	85.30 %	59.31 %	48.20 %	20.53 %	90.32 %	87.03 %
Number of Rules (Leafs)				6	17		

Source: *developed by the authors*

Testing the selected and stored model using the additional data set provided for this purpose, which is not included in the stored model. Despite the fact that Decision Tree has strong correlation coefficient (0.98), M5P MLA (0.91) is selected for testing the operability of the stored model. This is conditioned by the fact that the Decision Tree contains 17 leaves or rules, while M5P has only 6. With the help of stored M5P model, Table 11 shows the resulting prediction error and percentage, with less than a 1% inaccuracy in predicting soil moisture using leaf colour:

***Table 11**. Data obtained during the model's approbation*

Inst.	Actual	Predicted	Error	Error, %	Leaf	Inst.	Actual	Predicted	Error	Error, %	Leaf
1	20.61	20.584	-0.03	**-0.13%**	young	**14**	20.61	20.577	-0.03	**-0.16%**	young
2	17.85	17.747	-0.10	**-0.58%**	young	**15**	20.61	20.577	-0.03	**-0.16%**	old
3	17.85	17.973	0.12	**0.69%**	old	**16**	17.85	17.734	-0.12	**-0.65%**	young
4	20.61	20.577	-0.03	**-0.16%**	young	**17**	17.85	17.734	-0.12	**-0.65%**	young
5	17.85	17.975	0.13	**0.70%**	young	**18**	17.85	17.97	0.12	**0.67%**	old
6	20.61	20.577	-0.03	**-0.16%**	old	**19**	17.85	17.747	-0.10	**-0.58%**	young
7	17.85	17.747	-0.10	**-0.58%**	young	**20**	20.61	20.584	-0.03	**-0.13%**	young
8	20.61	20.584	-0.03	**-0.13%**	young	**21**	17.85	17.973	0.12	**0.69%**	old
9	17.85	17.973	0.12	**0.69%**	old	**22**	20.61	20.582	-0.03	**-0.14%**	young
10	17.85	17.984	0.13	**0.75%**	young	23	17.85	17.743	-0.11	**-0.60%**	young
11	17.85	17.751	-0.10	**-0.55%**	young	**24**	17.85	17.973	0.12	**0.69%**	old
12	17.85	17.984	0.13	**0.75%**	old	**25**	17.85	17.746	-0.10	**-0.58%**	young
13	17.85	17.975	0.13	**0.70%**	young	**26**	20.61	20.582	-0.03	**-0.14%**	young

***Source:** developed by the authors*

To determine the requirement for watering, it is feasible to develop and apply mathematical models that predict soil moisture content and plant water content, based on gathered data utilising conventional statistical approaches, using machine learning classification models (neural network, support vector machine, random forest and so on), combined – using classic methods and machine learning.

Guo *et al.* (2017) combined phenotyping and machine learning techniques to create a discrimination method for plant root zone water status in a greenhouse. Three root zone moisture levels, 40%, 60%, and 80% relative water content, were used and treated on pakchoi plants. Random Forest (RF), Neural Network (NN), and Support Vector Machine (SVM) classification models were developed and validated in different conditions, with overall accuracy of more than 90% for all. The SVM model had the highest value, but it required the most time to train. In all scenarios, all models achieved accuracy greater than 85%, while the RF model had more stable performance. The simplified SVM model created by the top five most contributing traits showed the greatest accuracy reduction of 29.5%, while the simplified RF and NN models maintained roughly 80% accuracy.

The study by Alhnaity *et al.* (2019) uses ML and DL algorithms to predict production and plant growth variance in two different scenarios, tomato yield forecasting and Ficus benjamina stem growth, under controlled greenhouse conditions. In the prediction formulas, they used a new deep recurrent neural network (RNN) model based on the Long Short-Term Memory (LSTM). The RNN architecture models the intended growth parameters using the previous yield, growth, and stem diameter values, as well as the microclimate variables. A comparison analysis is presented that employs ML approaches such as support vector regression and random forest regression, as well as the mean square error criterion, to evaluate the performance of the various methods.

Gutiérrez *et al.* (2018) describe a mobile method for estimating vineyard water status based on thermal imaging and machine learning. A thermal camera was mounted on an all-terrain vehicle traveling at 5 km/h to capture on-the-fly thermal photos of the vineyard canopy at 1.2 m and 1.0 m above ground. The canopy's two sides were measured in order to create side-specific and global models. The potential of stem water was measured and utilised as a reference method. Additionally, reference temperatures Tdry and Twet were determined for the calculation of two thermal indices: the crop water stress index (CWSI) and the Jones index (Ig). Prediction models were created both with and without the use of reference temperatures as input to the training algorithms. Using the reference temperatures, the best models had R^2 determination coefficients of 0.61 and 0.58 for cross validation and prediction, respectively (RMSE values of 0.190 MPa and 0.204 MPa). Yet, when the reference temperatures were not considered during model training, their performance statistics responded similarly, with R^2 values up to 0.62 and 0.65 for cross validation and prediction, respectively (RMSE values of 0.190 MPa and 0.184 MPa).

Dhillon (2015) created a continuous leaf monitoring system to monitor the water status of almond and walnut plants. The "Leaf Monitor" technology continuously measured leaf temperature and other microclimatic factors in the area of the leaf to check plant water status. It included a thermal infrared sensor to monitor leaf temperature, and sensors to measure ambient temperature and relative humidity, photosynthetically active radiation (PAR), and wind speed. The sensor system additionally included a leaf holder, a solar radiation diffuser dome, and a wind barrier to increase the unit's effectiveness. Each leaf monitor system was integrated into a mesh network of wireless nodes, allowing data collection and transmission over the web at 16-minute intervals.

Based on the foregoing approach, the same researcher and his team (Dhillon *et al.*, 2019) created a method for predicting plant water status in almond and walnut trees using a continuous leaf monitoring system. Using leaf temperature and environmental data collected by the leaf monitor, a Modified Crop Water Stress Index (MCWSI) was constructed to measure plant water status. A linear association with a coefficient of determination (r^2) of 0.67 was discovered in the instance of walnut crop. In the instance of almonds, a quadratic association with a coefficient of multiple determination (R^2) value of 0.75 was discovered. Based on the relationship discovered between MCWSI and DSWP, the irrigation amount for low frequency variable rate irrigation (VRI) was calculated, and variable rate irrigation (VRI) resulted in an average 39% reduction in water usage when compared to the fixed 100% ET replacement irrigation method for all trees. Based on the findings, the leaf monitor has the potential to be used as an irrigation scheduling tool.

Japanese scientists (Zhao *et al.*, 2020) investigated the application of hyperspectral imaging to assess the water status of tomato leaves in plant factories. Five varieties of tomatoes were sampled for testing in this experiment – three Japanese varieties suitable for direct consumption and two Dutch varieties suitable for processing and cooking. Three plants were selected from each of the five varieties. A portable hyperspectral camera was used to non-destructively measure the water status of tomato leaves. The measurement area was divided into three parts: leaves were measured from the highest part of the plants (near the growing point, 250 cm from the ground), from the middle of the plants (150 cm from the ground) and from the bottom of the plants (60 cm from the ground). Sampling was done 267 days after the plants were sown. 45 measurements (photographs) and samples were taken. After processing and analysing the hyperspectral data the tomato leaf raw relative reflection (RAW), inversion-logarithm relative reflection (LOG), and first derivative of relative reflection (DIFF) were analysed using the normalised difference vegetation index (NDVI) and two-band vegetation index (TBI) from 900 nm to 1700 nm. TBI regression using DIFF at wavelengths of 1.410 nm and 1.520 nm produced the best regression model for WC assessment, whereas NDVI regression using RAW produced the best regression model for MC evaluation. The MC evaluation produced better model performance than the WC assessment. The findings will aid our knowledge of the link between hyperspectral reflectance and leaf hydration status.

Except studies of Atanasov *et al.* (2016) – conventional statistical approach and Atanasov (2021) – using machine learning techniques, there is a lack of information regarding the colour of the leaves to be utilised as an indicator of the requirement for irrigation or any water stress. Difference between these previous studies and the current one is as follows: using nonlinear regression statistical Quasi-Newton method in (Atanasov *et al.*, 2016) soil moisture is predicted utilising piecewise linear regression and using data mining and machine learning techniques by Atanasov (2021). Both studies (Atanasov *et al.*, 2016) and (Atanasov, 2021) are conducted in different tomato varieties and different conditions of tomato growing (greenhouse).

There is a lack of scientific research that establishes a direct relationship or models (whether linear or non-linear) between plant colour change (lightening or darkening) and thus predicting soil moisture and the need for watering.

CONCLUSIONS

With the help of machine learning techniques and data mining, the relationship between RGB colour values from open field tomato leaves and soil moisture and temperature is investigated and modelled. It was proven nonlinear relationship. The name "12-9-6-3" for the methodology of measurements of field data was given. It is proven that the young leaves are more informative about the need for watering.

In the process of predicting soil moisture based on the colour of field tomato leaves and soil temperature, seven regression machine learning classifiers with three training data sets were trained: Zero Rule, Linear Regression, Locally Weighted Learning, M5P Model Tree, Support Vector Machin, Decision Tree and Multilayer Perceptron. For each MLA a classifier was trained, and then a model was created and saved. The efficiency of the chosen model was tested using a different test data set.

M5P produced best results. It was chosen to check the performance with M5P (0.91), despite that Decision Tree has a high correlation coefficient (0.98) since Decision Tree contains a significant number of rules (17), whereas M5P only has 6 (Table 10). Using the stored M5P model, Table 11 displays the resulting prediction error and percentage.

Based on the findings in Tables 3 to 9, it was demonstrated again that young leaves before watering are a stronger indicator of the need for irrigation (better results received than using the old leaves). As a result, the chosen regression model M5P predicts soil moisture with a precision of 1% error based on the colour of the tomato leaves and soil temperature. Such soil moisture prediction models, based on plant leaf colour change, besides saving labour time, automating routine repetitive actions, saving money and resources for expensive sensors and equipment, can be the basis of creating automated systems for automatic irrigation and saving water – only in the areas where it is needed, not in the entire agricultural field, a kind of precise irrigation based on the colour of the leaves of the plants themselves, the leaves serve as a natural biosensor.

ACKNOWLEDGEMENTS

This study was partially supported by the Bulgarian Ministry of Education and Science under the National Research Programme "Smart crop production" approved by Decision of the Ministry Council No. 866 / 26.11.2020.

CONFLICT OF INTEREST

The authors report no conflict of interest.

REFERENCES

[1] Afonso, Z.V., Claudinei, M.G., Elias, F.S., José, A.M.F., Weverton, P.R., Ícaro, A.S., Ricardo, B., Cláudio, R.M., & Eliemar, C. (2020). Using a crop water stress index based on a sap flow method to estimate water status in conilon coffee plants. *Agricultural Water Management*, 241, article number 106343. doi: 10.1016/j.agwat.2020.106343.

[2] Afzal, A., Duiker, S.W., & Watson, J.E. (2017). Leaf thickness to predict plant water status. *Biosystems Engineering*, 15, 148-156. doi: 10.1016/j.biosystemseng.2017.01.011.

[3] Alhnaity, B., Pearson, S., Leontidis, G., & Kollias, S. (2019). Using deep learning to predict plant growth and yield in greenhouse environments. *Acta Horticulturae*, 1296, 425-432. doi: 10.17660/ActaHortic.2020.1296.55.

[4] Atanasov, S. (2015). Soil specific FDR sensor calibration in soil moisture measuring (Bg). *Research Papers of Rousse University*, 54, 217-221.

[5] Atanasov, S. (2021). Predicting soil moisture based on color of the leaves using data mining and machine learning techniques. *IOP Conference Series: Materials Science and Engineering*, 1031, article number 012076. doi: 10.1088/1757-899X/1031/1/012076.

[6] Atanasov, S.S., Daskalov, P.I., & Nedeva, V.I. (2016). An intelligent approach of determining relationship between tomato leaves color and soil moisture and temperature. *Bulgarian Journal of Agricultural Science*, 22(6), 1027-1035.

[7] Bianchi, A., Masseroni, D., Thalheimer, M., de Medici, L.O., & Facchi, A. (2017). Field irrigation management through soil water potential measurements: A review. *Italian Journal of Agrometeorology*, 22(2), 25-38. doi: 10.19199/2017.2.2038-5625.025.

[8] Bouckaert, R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., & Scuse, D. (2019). *WEKA manual for version 3-8-4*. Hamilton: University of Waikato.

[9] Brownlee, J. (2016) *Machine learning mastery with Weka*. Retrieved from https://pdfcoffee.com/machine-learning-mastery-with-weka-2-pdf-free.html.

[10] Dhillon, R. (2015). *Development and evaluation of a continuous leaf monitoring system for measurement of plant water status*. (PhD dissertation, Department of Biological Systems Engineering, University of California, Davis).

[11] Dhillon, R., Rojo, F., Upadhyaya, S.K., Roach, J., Coates, R., & Delwiche, M. (2019). Prediction of plant water status in almond and walnut trees using a continuous leaf monitoring system. *Precision Agriculture*, 20, 723-745. doi: 10.1007/s11119-018-9607-0.

[12] Dobriyal, P., Qureshi, A., Badola, R., & Hussain, S.A. (2012). A review of the methods available for estimating soil moisture and its implications for water resource management. *Journal of Hydrology*, 458, 110-117. doi: 10.1016/j.jhydrol.2012.06.021.

[13] Eibe, F., Hall, M., & Witten, I.H. (2016). *The WEKA workbench. Online appendix for "Data mining: practical machine learning tools and techniques"*. Hamilton: The University of Waikato.

[14] Guo, D., Juan, J., Chang, L., Zhang, J., & Danfeng, H. (2017). Discrimination of plant root zone water status in greenhouse production based on phenotyping and machine learning techniques. *Scientific Reports*, 7(1), article number 8303. doi: 10.1038/s41598-017-08235-z.

[15] Gutiérrez, S., Diago, M.P., Fernández-Novales, J., & Tardaguila, J. (2018). Vineyard water status assessment using on-the-go thermal imaging and machine learning. *PloS One*, 13(2), article number e0192037. doi: 10.1371/journal.pone.0192037.

[16] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I.H. (2009). The WEKA Data Mining Software: An Update. *ACM SIGKDD Explorations Newsletter*, 11(1), 10-18. doi: 10.1145/1656274.1656278.

[17] Harizanova-Petrova, B., & Ovcharova, A. (2014). Investigation the relationship "Yield–Evapotranspiration" by stages of rooted celery, variety "IBIS" in the region of Plovdiv. *Turkish Journal of Agricultural and Natural Sciences*, 1(1), 875-880.

[18] Huan, Z., Liangxiao, J., & Liangjun, Y. (2021). Attribute and instance weighted naive Bayes. *Pattern Recognition*, 111, article number 107674. doi: 10.1016/j.patcog.2020.107674.

[19] Jägermeyr, J., Gerten, D., Heinke, J., Schaphoff, S., Kummu, M., & Lucht, W. (2015). Water savings potentials of irrigation systems: Global simulation of processes and linkages. *Hydrology and Earth System Sciences*, 19(7), 3073-3091. doi: 10.5194/hess-19-3073-2015.
[20] Kalaidjieva, R., Kuneva, V., & Matev, A. (2015). Using on infrared thermometer for establishment of water stress in green beans irrigated by different pre-irrigation soil moisture. *Plant Science*, 5, 72-78.
[21] Kocaoglu, S., & Akdogan, E. (2020). Comparison of classification algorithms for detecting patient posture in expandable tumor prostheses. *Advances in Electrical and Computer Engineering*, 20(2), 131-138. doi: 10.4316/AECE.2020.02015.
[22] Ko-Madden, C.T., Upadhyaya, S.K., Kizer, E.E., Drechsler, K.M., Rojo, F., Meyers, J.N., & Schramm, A.E. (2017). *Precision irrigation in wine grape using a proximal leaf monitor system for measuring plant water status*. Michigan: American Society of Agricultural and Biological Engineers. doi: 10.13031/aim.201700705.
[23] Koumanov, K., Matev, A., Kornov. G., & Germanova, D. (2018). Irrigation management in perennial crops by sensing the soil water status. *Journal of Mountain Agriculture on the Balkans*, 21(5), 182-192.
[24] Levin, A.D. (2019). Re-evaluating pressure chamber methods of water status determination in field-grown grapevine (Vitis spp.). *Agricultural Water Management*, 221, 422-429. doi: 10.1016/j.agwat.2019.03.026.
[25] Mahan, J.R., Young, A.W., & Payton, P. (2015). Continuously monitored canopy temperature as a proxy for plant water status. *American Journal of Plant Sciences*, 6(14), article number 2287. doi: 10.4236/ajps.2015.614232.
[26] Nederhoff, E., & Stanghellini, C. (2010). Water use efficiency of tomatoes. *Practical Hydroponics and Greenhouses*, 115, 52-59.
[27] Pardossi, A., Incrocci, L., Incrocci, G., Malorgio, F., Battista, P., Bacci, L., & Balendonck, J. (2009). Root zone sensors for irrigation management in intensive agriculture. *Sensors*, 9(4), 2809-2835. doi: 10.3390/s90402809.
[28] Sela, E., Cohen, Y., Alchanatis, V., Saranga, Y., Cohen, S., Möller, M., & Orolov, V. (2007). Thermal imaging for estimating and mapping crop water stress in cotton. Wageningen: Wageningen Academic Publishers.
[29] Smola, A.J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199-222.
[30] United Nations Development Programme, Precision agriculture for smallholder farmers. (2021). Retrieved from https://www.undp.org/publications/precision-agriculture-smallholder-farmers.
[31] Witten, I.H. (2013). *Data mining with Weka*. Waikato: University of Waikato.
[32] Zhao, T., Nakano, A., Iwaski, Y., & Umeda, H. (2020). Application of hyperspectral imaging for assessment of tomato leaf water status in plant factories. *Applied Sciences*, 10(13), article number 4665. doi: 10.3390/app10134665.

Колір листя томатів як показник вологості ґрунту з використанням методів машинного навчання

Светослав Атанасов

Аспірант, магістр комп'ютерних систем і технологій
Фракійський університет
6015, Студентське містечко, м. Стара Загора, Болгарія
https://orcid.org/0000-0002-2658-1611

Біляна Харізанова-Петрова

Доктор філософії, головний доцент
Аграрний університет
4000, бульвар Менделєєва, 12, м. Пловдив, Болгарія
https://orcid.org/0000-0001-8437-7718

Радость Петрова

Кандидат технічних наук, доцент
Аграрний університет
4000, бульвар Менделєєва, 12, м. Пловдив, Болгарія
https://orcid.org/0000-0002-4476-7049

Анотація. Запаси прісної води для зрошення повинні використовуватися економно і розумно, оскільки вода є безцінним природним ресурсом, якого не вистачає на більшій частині Землі. Вологість ґрунту на полях не всюди однакова, а встановлення тисяч датчиків є невиправдано дорогим. Мета цієї публікації - змоделювати та спрогнозувати взаємозв'язок між кольором листя рослин томатів та вологістю ґрунту, і таким чином оптимально керувати процесом зрошення. Дослідження проводили з використанням загальноприйнятих методів, польового методу та методу статистичної обробки результатів. Алгоритми машинного навчання (АМН) та інтелектуального аналізу даних були використані в даній роботі для моделювання зв'язку між значеннями кольору RGB листя томатів та вологістю і температурою ґрунту. У фокусі цього дослідження – колір листя томатів, вирощених у відкритому ґрунті без кілків. Було виконано три основні завдання: доведено, що існує зв'язок між кольором листя і вологістю ґрунту, досліджено його ймовірний нелінійний тип і змодельовано цей зв'язок за допомогою АМН. Спочатку навчався класифікатор, потім створюється і зберігається модель. Нарешті, ефективність обраної моделі перевірено за допомогою іншого тестового набору даних. Методологія вимірювань отримала назву "12-9-6-3". Доведено, що молоде листя є більш інформативним щодо потреби в поливі. В результаті, за допомогою регресійної моделі M5P можна, з похибкою менше 1%, прогнозувати вологість ґрунту за кольором листя томатів, враховуючи також температуру ґрунту. Ця прогнозна модель може бути використана при створенні автоматизованих систем для оптимального управління поливом та водозбереження

Ключові слова: водний статус рослин; вологість ґрунту; колір листя; колір накриття; нелінійне оцінювання