Non-Destructive Determination of Plant Pigments Based on Mobile Phone Data

Miroslav Vasilev [1], Vanya Stoykova [1], Petya Veleva [2], Zlatin Zlatev [1]

[1]*Faculty of technics and technologies, Trakia University, 38 Graf Ignatiev str., 8602, Yambol, Bulgaria*
[2]*Department of Agricultural Engineering, Faculty of Agriculture, Trakia University, Student's campus, 6000 Stara Zagora, Bulgaria*

***Abstract* – This paper proposes methods and tools for determining plant pigments using data from a mobile phone video sensor. A disadvantage of the known studies in this field is that they are mainly aimed at determining the chlorophyll content. There are few studies related to the determination of pigments such as carotenoids, flavonoids, and betalains, which are also important in terms of determining the condition of plants in their cultivation. Cucumbers were chosen because the long periods of drought in Bulgaria, lead to losses in cultivating these plants. Vectors containing colour and spectral indices were used. These features are obtained through a video sensor on a mobile phone. The kernel method variant of principal components reduces them. Feature vectors are selected using factor analysis, correspondence analysis, and the correlation method. Predictive models have been developed to determine plant pigments. With high accuracy (over 90%), the pigments xanthophyll and chlorophyll A can be predicted.**

***Keywords* – Visual analysis, spectral data, mobile phone video sensor, leaves coloration, carotenoids, chlorophyll content.**

DOI: 10.18421/TEM123-23
https://doi.org/10.18421/TEM123-23

Corresponding author: Petya Veleva,
Department of Agricultural Engineering, Faculty of Agriculture, Trakia University, Student's campus, 6000 Stara Zagora, Bulgaria
Email: petya.veleva@trakia-uni.bg

Received: 24 April 2023.
Revised: 06 June 2023.
Accepted: 04 August 2023.
Published: 28 August 2023.

1. Introduction

Pigments are indispensable components of plants, enabling them to harness light energy, protect themselves against environmental stress, and perform essential functions like photosynthesis and reproduction.

Understanding plant pigments and their roles in growth and adaptation is crucial for optimizing agricultural practices, maintaining ecosystem health, and ensuring sustainable plant life. The main plant pigments include carotenoids, flavonoids, betalains, and chlorophylls. They are chemical compounds that absorb certain wavelengths of visible light. The light is reflected by the structures containing the pigments [1].

Carotenoids are a group of natural tetraterpenoid pigments. Xanthophylls play a vital role in plant photosynthesis and photoprotection. They are key components of the plant's photosynthetic apparatus, connecting the photosystem, aiding in the light collection, and providing photoprotection. It is an important protective pigment of the photosynthetic apparatus of plants.

The content of chlorophyll in the leaves of plants provides information about their physiological status, such as nutrient deficiencies. It is directly related to the photosynthetic potential of plants and is an important indicator of their cultivation. Chlorophyll A absorbs light in the blue, red, and violet spectrum, playing a crucial role in oxygen photosynthesis. Chlorophyll B is primarily responsible for absorbing blue light during photosynthesis. The amount of chlorophyll in plants is influenced by factors like nitrogen content and water stress, which can negatively impact plant development [2].

Classical laboratory methods for determining the content of pigments in plants such as chromatographic and spectrophotometric are time-consuming, and require high-cost equipment and experienced operators.

In precise agriculture, there are used computer systems based on computer vision [22] or spectral data analysis [23], [24] for the assessment of quality properties of different agricultural products.

Table 1 presents a comparative analysis of the efficient methods developed for the express determination of pigments in plants, considering the current level of scientific and technological advancements.

The comparative analysis of non-destructive methods for determining pigment levels in plants revealed a predominant focus on quantifying chlorophyll content.

Insufficient research hinders understanding essential pigments like carotenoids, flavonoids, betalains, etc., and their significant role in assessing plant health and cultivation.

Table 1. Advantages and disadvantages of methods for non-destructive determination of plant pigments

Method	Advantages	Disadvantages	Reference
Video camera	Covers an area of the object	Influence of measurement conditions	[3]
Colorimeter	Less influence of measurement conditions	Receive data at one point of the object	[4]
Spectrophotometer	Obtaining spectral characteristics	Receive data at one point of the object	[5]
Hyperspectral analysis	Covers an area of the object. Obtaining spectral characteristics	Relatively expensive technology	[6]
Mobile phone	Affordable device. Covers an area of the object	Influence of measurement conditions	[7]

A common disadvantage of the colorimeter, video camera, and mobile phone measurement methods is that the available publications mainly use data from the RGB colour model, which is hardware dependent, and the values obtained are largely influenced by the conditions of obtaining digital images.

In methods based on spectral analysis, commercial devices for determining the chlorophyll content directly in the field are available. A modern trend is to apply methods of artificial intelligence and machine learning to process the obtained spectral characteristics. These methods greatly enhance the accuracy of chlorophyll content determination in plants, surpassing classical regression analysis. However, a drawback is the need for post-processing of computer data, which hinders direct field assessment of plant conditions.

Systems utilizing neural networks have the drawback of relying on complex computational procedures, making them more suitable for laboratory settings. The analysis indicates that classification methods utilizing simplified calculation procedures are more suitable, as they can be applied in field-based technical equipment. These methods expedite calculations and reduce system response time.

In recent years, solutions using mobile phones have become more common in the field of determining the pigment content of plants. This is because these devices are affordable, and their video sensors have good characteristics to obtain output data for the captured objects.

As mentioned, a major drawback is that data is used directly from the RGB colour model and the measurement is greatly influenced by the image acquisition conditions. There are solutions with the addition of spectrophotometric sensor devices that connect to a mobile phone [7]. The primary drawback of such devices is the significant increase in cost. Research using a mobile phone is mainly focused on the determination of chlorophyll in plants. There are few publications in which other pigments such as carotenes, xanthophyll, flavonoids, betalains, and others have been identified.

Long periods of drought leading to losses in the cultivation of cucumbers are common in Bulgaria. Proper monitoring of drought stages is crucial to prevent disturbances in plant development. This necessitates the search for methods by which to signal in a timely manner changes in the normal state of these plants.

The purpose of this work is to propose methods and tools for determining the pigments of cucumbers using data from a video sensor, integrated into a mobile phone. A simple and affordable method of non-invasive pigment evaluation would be valuable to plant biologists.

2. Material and Methods

Cucumbers of the variety “Gergana” (Floriyan Ltd, Bulgaria) were used. They were grown under adapted conditions. They were sown in containers with a volume of 1000 cm^3.

The soil in which they were sown was from the producer Semenata Shop, Bulgaria. The irrigation rate was 150 ml every 4 days.

Cucumbers were divided into two groups: 5-6 leaves of their development and in the phase of active flowering. The subject of the study was the change in the content of cucumber pigments under different soil moisture conditions.

The soil moisture was measured by KIT K2112 (Pasat Electronics OOD, Gorna Oryahovitsa, Bulgaria), representing a two-electrode measuring system, with a 9V DC power supply. There is a possibility to measure soil moisture using a microprocessor system based on image analysis [25].

A modified version of Priyadarshini et al. [8] method to determine pigments was employed. Cucumber leaves, trimmed to 10x10 mm pieces, were soaked in acetone for 4 hours after removing the stems. A drop of the extract was applied onto a 105x19 mm white paper with a density of 80 g/m^2, positioned at a distance of 15 mm. The sample was air-dried for 15 minutes. Subsequently, 4 mm of the paper end was dipped in acetone. After 15 minutes, the samples were removed from the acetone and left to dry for an hour. The Rf values of four separate fractions were recorded, with three repetitions conducted to determine the mean and standard deviation. Rf pigment values were calculated as a percentage of the total pigment concentration, using the formula Rf=(A/B)x100%, where A represents the distance travelled by the solvent, and B represents the distance travelled by the respective fraction. The reported fractions were Carotene (C), Xanthophyll (X), Chlorophyll A (HA), and Chlorophyll B (HB).

Samsung's 48 MP ISOCELL Bright GM2 video sensor was used to obtain colour digital images. It has a 1/2,25 sensor with a pixel size of 0,8 µm, and the lens has an f/1,8 aperture. This sensor is integrated into a Xiaomi Redmi Note 8T mobile phone.

Measurements were taken over 12 days. 30 images of each type of cucumber were taken per day. A total of 720 color digital images of the cucumbers were recorded for the entire measurement period.

The distance between the camera and the object was 25 cm. A homogeneous illumination source was used, consisting of white LEDs with a color temperature of 6400 K and the greatest intensity of emitted light at 450 nm.

The R, G, and B matrices of the images were transformed into component vectors. The RGB color model components (RGB [0 255]) were converted to Lab (L [0 100], a [-86.18 98.23], b [-107.86 94.47]) using an online conversion tool called Convert Rgb to Lab (colormine.org). The chroma (C) and hue values from the LCh color model [9] were determined using the following method:

$$C = \sqrt{a + b} \qquad h^o = atan\left(\frac{b}{a}\right) \tag{1}$$

The indices of yellow, white, and brown are determined, as well as those according to the formulas summarized in the available literature. The L, a, and b components from the Lab color model were used. Also, C and h from the LCh model [10], [11]:

$$YI = \frac{142,86b}{L} \tag{2}$$

$$WI = 100 - \sqrt{(100 - L)^2 + a^2 + b^2} \tag{3}$$

$$BI = \frac{x - 0,31}{0,17}, \text{ where } x = \frac{a + 1,75L}{5,645L + a - 0,012b} \tag{4}$$

$$SI = \sqrt{a^2 + b^2} \tag{5}$$

$$CIRG = \frac{180 - h}{L + C} \tag{6}$$

$$COL = \frac{2000a}{LC} \tag{7}$$

$$CI = \frac{a}{b} \tag{8}$$

$$ECB = \frac{a}{b} + \frac{a}{L} \tag{9}$$

$$FCI = L - b \tag{10}$$

$$WL = \frac{L}{b} \tag{11}$$

$$PACI = \frac{1000a}{L + h} \tag{12}$$

The conversion of values between the XYZ and LMS models to reflection spectra in the visible (VIS) region (390-730nm) was achieved using mathematical relationships that allow conversion in both directions [12]. The conversion process utilized data based on observer 2° and illumination D65 (average daylight with a UV component, 6500K). The conversion of VIS spectral characteristics employed spectral indices defined by Cermakova et al. [13] and Atanassova et al. [14]. These indices capture specific parts of the light spectrum, including red, green, blue, and yellow-orange regions. The indices were calculated based on reflection spectra (R) using the following formulas:

$$REI = \frac{R_{740}}{R_{720}} \quad (13)$$

$$PTI = \frac{R_{530} - R_{570}}{R_{530} + R_{570}} \quad (14)$$

$$CTI = \frac{1}{R_{510}} - \frac{1}{R_{550}} \quad (15)$$

$$TVI = 0{,}5(120(R_{750} - R_{550}) - 200(R_{670} - R_{550})) \quad (16)$$

$$G = \frac{R_{550}}{R_{680}} \quad (17)$$

$$NExG = \frac{2R_{520} - R_{620} - R_{420}}{R_{520} + R_{620} + R_{420}} \quad (18)$$

$$NGRDI = \frac{R_{520} - R_{620}}{R_{520} + R_{620}} \quad (19)$$

$$RGBVI = \frac{{R_{520}}^2 - R_{620}R_{420}}{{R_{520}}^2 + R_{620}R_{420}} \quad (20)$$

$$GLI = \frac{2R_{520} - R_{620} - R_{420}}{2R_{520} + R_{620} + R_{420}} \quad (21)$$

$$VARI = \frac{R_{520} - R_{620}}{R_{520} + R_{620} - R_{420}} \quad (22)$$

$$ExG = 2R_{520} - R_{620} - R_{420} \quad (23)$$

Informative feature selection was performed using factor analysis (FA), correspondence analysis (CA), and the correlation method (Corr).

In *factor analysis* (FA), the maximum similarity of factor weights was computed in a matrix Λ, based on the factor analysis model [15].

$$x = \mu + \Lambda f + e \quad (24)$$

where x is a vector of observations; μ - vector of the average values; Λ - matrix with dimensions dxm of factor weights; f - vector of independent, standardized common factors; e - vector of independent specific factors. X, μ, and e have dimension d, and f has dimension m.

Correspondence analysis (CA) is a method used to analyze relationships between two sets of categorical data [16]. It is suitable for matrices where the elements represent the frequencies of co-occurring events for different categories of the two factors, represented by rows and columns in a table. The analysis yields results in the form of vectors r_i and c_j. The weights for rows (w_i) and columns (w_j) are derived from these vectors:

$$w_i = \{r_i\} \qquad w_j = \{c_j\} \quad (25)$$

Correlation method (Corr). The correlation analysis enables the exploration of potential relationships between the features that describe variations in pigments. During feature selection, the focus is on identifying those features that exhibit a correlation with the measurement time (t, h) greater than 0,6.

The key criterion for selecting characteristics to form a feature vector involved considering only those with a weight coefficient above 0,6 in at least two of the selection methods.

The selected features were structured into feature vectors, and the data within these vectors were reduced using the kernel variant of the Principal Components Analysis method (kPCA) [17], [18]. Three different kernel functions were employed: Simple (Linear), Polynomial, and Gaussian. An initial model was utilized to capture the relationship between the selected characteristics of cucumbers, such as:

$$z = b_0 + b_1 x + b_2 y + b_3 x^2 + b_4 xy + b_5 y^2 \quad (26)$$

The accuracy of the predictive models was assessed using statistical parameters such as the coefficient of determination (R^2) and standard errors. These parameters allowed for the analysis and interpretation of empirical data [19]. The model coefficients, their corresponding standard errors (SE), t-statistics (tStat), and p-values were calculated. Additionally, the F- criterion was employed, where the calculated F-value needed to exceed the critical F-value. The coefficient analysis relied on the p-value, allowing the identification and exclusion of non-informative coefficients from the model. Residual analysis was conducted by calculating the differences between the model-predicted values and the actual measurements, resulting in the residuals ($r_i = y - y_{fit}$).

In the present work, the software products Matlab (The Math Works Inc.), Stat Soft Statistica 12 (Stat Soft Inc.), and MS Office 2016 (Microsoft Corp.) were used. All data were processed at a level of significance α = 0,05.

3. Results and Discussion

A study was conducted to compare the RGB values and spectral characteristics of cucumbers with and without watering. Additionally, models were developed to analyze the pigment features of cucumbers in both watered and non-watered conditions.

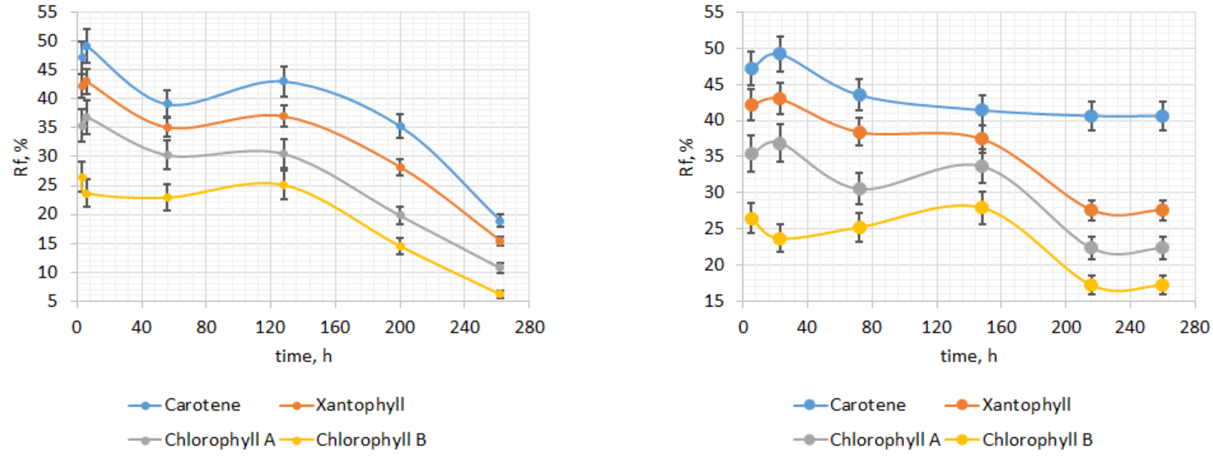
3.1. The RGB Values and Spectral Characteristics of Cucumbers with and Without Watering

In cucumbers without watering, the amount of pigments decreases, while in those with watering, this amount varies depending on the watering regime.

In cucumbers without watering, the colour changes from lighter to darker during the measurement period. This is due to the reduction in the amount of the pigment chlorophyll in the leaves due to drought.

In cucumbers with watering, the colour changes in proportion to the watering regime. At the beginning of the drying process, the colour starts to transition towards a darker shade, and after watering, it undergoes a reversal and shifts towards a lighter hue once again.

Figure 1 shows the change in the Rf, % values of the pigments in the variants with and without watering.



a) without watering *b) with watering*

Figure 1. Change in the values of the cucumber pigments

Figure 2 illustrates the overall representation of average RGB values for cucumbers in the absence of watering.

Figure 3 provides a comprehensive depiction of the average RGB values for cucumbers with watering.

Figure 4 shows the change in averaged spectral characteristics.

The change in spectral characteristics is similar to that of the colour components of the RGB model.

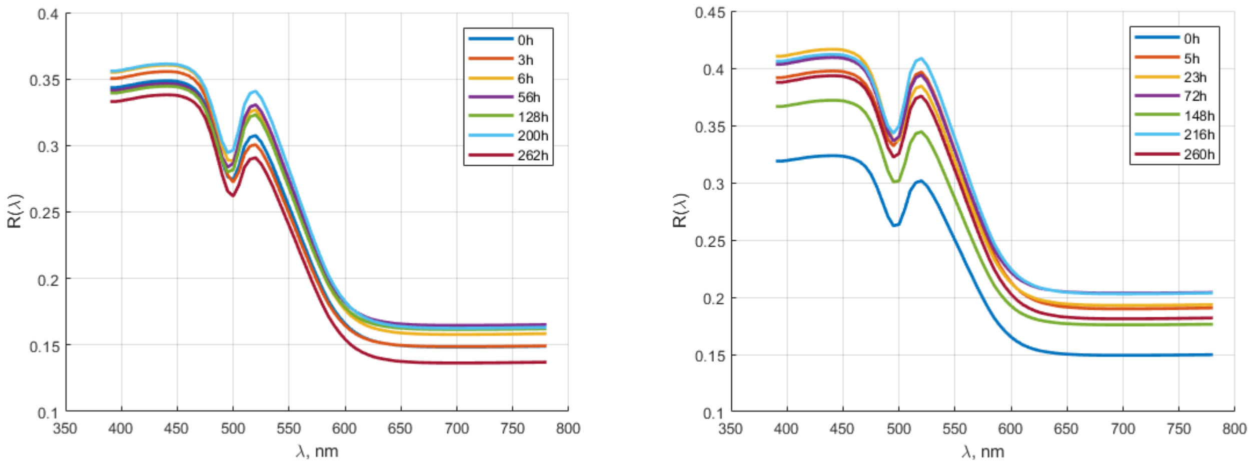
After the cucumbers are watered, the reflection values gradually increase, at all wavelengths. It can be seen that in all cases the peak values of the spectral characteristics are in the regions around 420 and 520 nm. In the visible range, these are the blue and green areas of the light spectrum.

0 h [56 106 51] | 3 h [62 109 47] | 6 h [57 109 56] | 56 h [62 104 56] | 128 h [63 104 54] | 200 h [54 109 61] | 262 h [48 103 49]

Figure 2. Averaged RGB values for cucumbers without watering – general view

0 h [56 98 51] | 5 h [59 119 73] | 23 h [77 126 63] | 72 h [84 123 64] | 148 h [73 112 55] | 216 h [73 123 72] | 260 h [62 118 67]

Figure 3. Averaged RGB values of cucumbers with watering – general view



a) without watering *b) with watering*

Figure 4. Change in averaged spectral characteristics

3.2. Modelling of the Features of the Pigments of Cucumbers Without Watering

Figure 5 displays the outcomes of feature selection for cucumbers in the absence of watering. The selection process involved utilizing the methods of Factor Analysis (FA), Correspondence Analysis (CA), and Correlation (Corr). Notably, a majority of the spectral indices chosen through the FA method exhibited weight coefficients exceeding 0.6. Furthermore, the color indices demonstrated weight coefficients surpassing 0,6 in at least two of the selection methods employed.

After using the criterion, at least by two of the selection methods, the corresponding feature to have weight coefficients of more than 0,6 the following vector of features for cucumbers without watering is formed:

$$\text{NIFV=[a C WI BI SI COL CI ECB FCI WL PACI PTI TVI G]} \tag{27}$$

One colour component from the Lab and LCh colour models, nine colour indices, and three spectral indices were selected. These results show that colour indices are more informative than spectral indices.

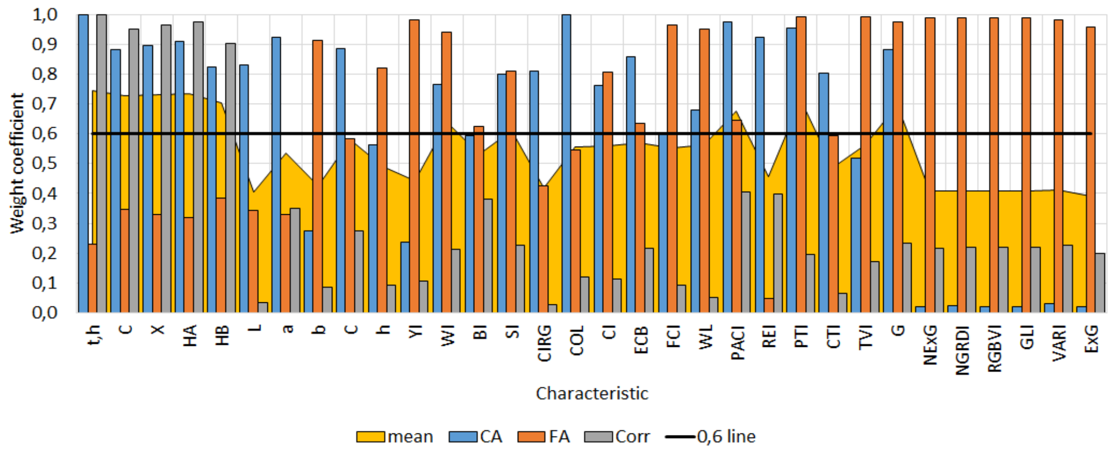


Figure 5. Results of selection of features for cucumbers without watering

The ability to predict cucumber pigments by reducing the resulting feature vector with kPCA with Simple Kernel was tested. Upon eliminating the non-significant coefficients from the primary models, it was determined that the association between the dependent variables (pigments) and the independent variables (the first two principal components) can be expressed by the following models:

$C=f(PC_1,PC_2)$
$$C = 0{,}38 - 0{,}23PC_2 - 0{,}17{PC_2}^2 + 0{,}02PC_1PC_2 \tag{28}$$

$X=f(PC_1,PC_2)$
$$X = 0{,}33 - 0{,}21PC_2 - 0{,}13{PC_2}^2 + 0{,}04PC_1PC_2 \tag{29}$$

$HA=f(PC_1,PC_2)$
$$HA = 0{,}25 - 0{,}003PC_1 - 0{,}25PC_2 - 0{,}13{PC_2}^2 \tag{30}$$

$HB=f(PC_1,PC_2)$
$$HB = 0{,}2 - 0{,}17PC_2 - 0{,}16{PC_2}^2 + 0{,}04PC_1PC_2 \tag{31}$$

Table 2 shows the values for estimating the parameters of the obtained models. It can be seen that the model for forecasting chlorophyll A is with the highest coefficient of determination and the lowest errors, indicating a strong relationship between the dependent variables (pigments) and the independent variables (the first two principal components). The value of the critical coefficient is Fcr (3,32) = 2,9.

Table 2. Evaluation of the obtained models (kPCA with Simple Kernel) for pigments of cucumbers without watering

Model	R^2	F(3,32)	SE	p
$C=f(PC_1,PC_2)$	0,95	209,47	0,02	<0,00
$X=f(PC_1,PC_2)$	0,95	210,79	0,02	<0,00
$HA=f(PC_1,PC_2)$	0,94	180,19	0,02	<0,00
$HB=f(PC_1,PC_2)$	0,94	177,04	0,02	<0,00

The ability to predict cucumber pigments by reducing the resulting feature vector by kPCA with a Polynomial kernel was tested. After removing the insignificant coefficients from the main models that have p-value >> α, it was found that the relationship between the dependent variables – pigments and the independent variables, can be described by the following models:

$C=f(PC_1,PC_2)$ $$C = 0,39 - 0,05PC_2 - 0,01{PC_2}^2 + 0,002PC_1PC_2 \quad (32)$$
$X=f(PC_1,PC_2)$ $$X = 0,34 - 0,05PC_2 - 0,01{PC_2}^2 + 0,002PC_1PC_2 \quad (33)$$
$HA=f(PC_1,PC_2)$ $$HA = 0,26 - 0,05PC_2 - 0,01{PC_2}^2 + 0,002PC_1PC_2 \quad (34)$$
$HB=f(PC_1,PC_2)$ $$HB = 0,21 - 0,04PC_2 - 0,01{PC_2}^2 + 0,002PC_1PC_2 \quad (35)$$

Table 3 presents the parameter estimates for the obtained models. Similar to the previous models, it is evident that the model for predicting chlorophyll A achieves the highest coefficient of determination and the lowest error values. Additionally, the critical coefficient value, Fcr (3,32) = 2,9, is also noteworthy.

Table 3. Evaluation of the obtained models (kPCA with a Polynomial kernel) for pigments of cucumbers without watering

Model	R^2	F(3,32)	SE	p
$C=f(PC_1,PC_2)$	0,96	272,5	0,02	<0,00
$X=f(PC_1,PC_2)$	0,96	272,15	0,02	<0,00
$HA=f(PC_1,PC_2)$	0,96	286,07	0,02	<0,00
$HB=f(PC_1,PC_2)$	0,96	253,47	0,01	<0,00

The ability to predict cucumber pigments by reducing the resulting feature vector by kPCA with a Gaussian kernel was tested. After removing the insignificant coefficients from the main models that have p-value >> α, it was found that the relationship between the dependent variables – pigments and the independent variables, can be described by the following models:

$C=f(PC_1,PC_2)$ $$C = 0,44 - 006PC_1 - 0,85PC_2 - 3,21{PC_2}^2 + 1,12PC_1PC_2 \quad (36)$$
$X=f(PC_1,PC_2)$ $$X = 0,39 - 0,05PC_1 - 0,77PC_2 - 2,57{PC_2}^2 + 0,9PC_1PC_2 \quad (37)$$
$HA=f(PC_1,PC_2)$ $$HA = 0,32 - 0,06PC_1 - 0,94PC_2 - 2,56{PC_2}^2 + 0,99PC_1PC_2 \quad (38)$$
$HB=f(PC_1,PC_2)$ $$HB = 0,25 - 0,04PC_1 - 0,53PC_2 - 3,04{PC_2}^2 + 0,71PC_1PC_2 \quad (39)$$

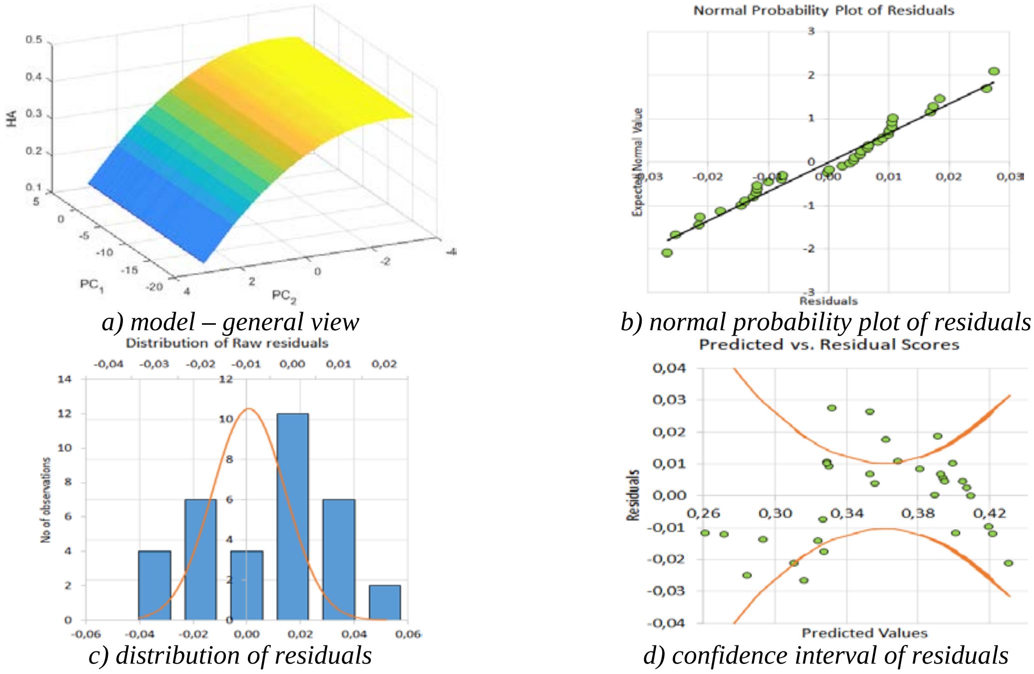
Table 4 displays the parameter estimates for the obtained models. Notably, the model for predicting chlorophyll A exhibits the highest coefficient of determination and the lowest error values. Moreover, the critical coefficient value is observed as Fcr (3,32) = 2,9.

Table 4. Evaluation of the obtained models (kPCA with a Gaussian kernel) for pigments of cucumbers without watering

Model	R^2	F(3,32)	SE	p
$C=f(PC_1,PC_2)$	0,96	174,42	0,02	<0,00
$X=f(PC_1,PC_2)$	0,96	175,26	0,02	<0,00
$HA=f(PC_1,PC_2)$	0,96	198,43	0,02	<0,00
$HB=f(PC_1,PC_2)$	0,95	143,94	0,01	<0,00

From the analysis of cucumber pigments performed during watering, it was found that chlorophyll A can be predicted with the greatest accuracy by the vector of features reduced by principal components with a polynomial kernel. The regression model has the form of equation (34).

Figure 6 shows the resulting model. Residual analysis was also performed.



a) model – general view *b) normal probability plot of residuals*

c) distribution of residuals *d) confidence interval of residuals*

Figure 6. Residual analysis of model HA = f (PC1, PC2) obtained by kPCA-polynomial kernel

The values of the first principal component are represented on the "x" axis of the model, while the values of the second principal component are shown on the "y" axis. The chlorophyll A (HA) values are plotted along the vertical axis.

Observing the distribution of the residuals and their proximity to the normal line, as well as considering the normal probability graph, it can be inferred that the residuals closely follow a normal distribution. Based on this criterion, it can be concluded that the prerequisites for regression analysis are met.

3.3. Modelling of the Features of the Pigments of Cucumbers with Watering

Figure 7 shows the results of the selection of features for cucumbers with watering. The selection of features was made by the methods of FA, CA, and Corr. It can be seen that the weight coefficients of the spectral indices have higher values than those of the colour indices.

After the criterion has been used, by at least two of the selection methods, the corresponding feature to have weight coefficients of more than 0,6 is formed in the following vector of features for cucumbers with watering:

$$IFV=[YI\ WI\ SI\ CI\ FCI\ WL\ PACI\ PTI\ TVI\ G\ NExG\ NGRDI\ RGBVI\ GLI\ VARI] \quad (40)$$

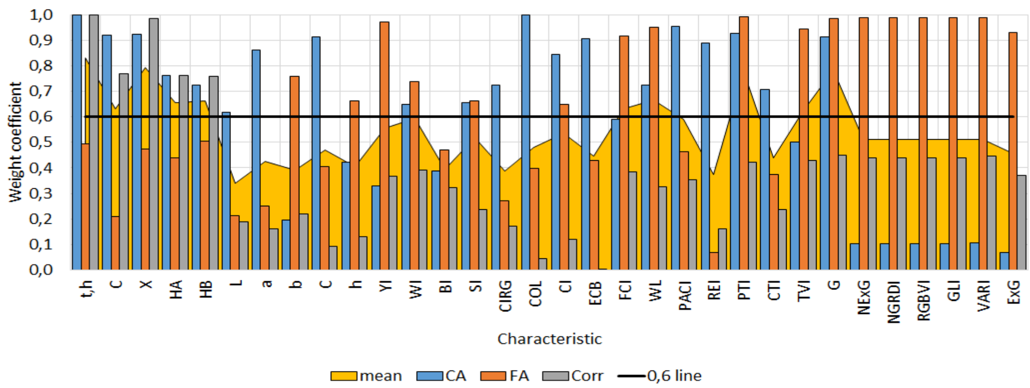


Figure 7. Results of selection of features for cucumbers with watering

Seven colour and eight spectral indices were selected. In this case, the colour and spectral indices are equally informative.

The ability to predict cucumber pigments by reducing the resulting feature vector by kPCA with Simple Kernel was tested.

Upon eliminating the non-significant coefficients from the main models with p-values significantly greater than α, it has been determined that the relationship between the dependent variables (pigments) and the independent variables can be accurately described by the following models:

$C=f(PC_1,PC_2)$	$C = 0{,}41 + 0{,}01{PC_1}^2 + 0{,}1PC_1PC_2$	(41)
$X=f(PC_1,PC_2)$	$X = 0{,}32 - 0{,}06PC_1 - 0{,}09PC_2 - 0{,}01{PC_1}^2$	(42)
$HA=f(PC_1,PC_2)$	$HA = 0{,}28 - 0{,}05PC_1 - 0{,}01{PC_1}^2 - 0{,}05PC_1PC_2$	(43)
$HB=f(PC_1,PC_2)$	$X = 0{,}21 - 0{,}04PC_1 - 0{,}04PC_2 - 0{,}01{PC_1}^2$	(44)

Table 5 presents the parameter estimates for the obtained models. It is evident that the model with the highest coefficient of determination is observed indicating a strong relationship between the variables and the lowest value of errors has the model for predicting xanthophyll. The value of the critical coefficient is Fcr (3,28) = 2,95.

Table 5. Evaluation of the obtained models (kPCA with Simple Kernel) for pigments of cucumbers with watering

Model	R^2	F(3,28)	SE	p
C=f(PC1,PC2)	0,64	25,43	0,02	<0,00
X=f(PC1,PC2)	0,89	76,56	0,02	<0,00
HA=f(PC1,PC2)	0,65	17,72	0,03	<0,00
HB=f(PC1,PC2)	0,67	19,08	0,02	<0,00

The ability to predict cucumber pigments by reducing the resulting feature vector by kPCA with a Polynomial kernel was tested. After removing the insignificant coefficients from the main models that have p-value >> α, it was found that the relationship between the dependent variables – pigments and the independent variables, can be described by the following models:

$C=f(PC_1,PC_2)$	$C = 0{,}51 - 0{,}18PC_2 - 0{,}0002{PC_1}^2 + 0{,}02PC_1PC_2$	(45)
$X=f(PC_1,PC_2)$	$X = 0{,}49 - 0{,}26PC_2 - 0{,}001{PC_1}^2 + 0{,}02PC_1PC_2$	(46)
$HA=f(PC_1,PC_2)$	$HA = 0{,}34 - 0{,}01PC_1 - 0{,}01{PC_2}^2$	(47)
$HB=f(PC_1,PC_2)$	$HB = 0{,}26 - 0{,}004PC_1 - 0{,}01{PC_2}^2$	(48)

Table 6 shows the values for estimating the parameters of the obtained models. It can be seen that the model for predicting xanthophyll has the highest value of the coefficient of determination and the lowest value of errors. The value of the critical coefficient is Fcr (3,28) = 2,95.

Table 6. Evaluation of the obtained models (kPCA with a Polynomial kernel) for pigments of cucumbers with watering

Model	R^2	F(3,28)	SE	p
C=f(PC1,PC2)	0,72	24,48	0,02	<0,00
X=f(PC1,PC2)	0,93	120,58	0,01	<0,00
HA=f(PC1,PC2)	0,64	25,57	0,03	<0,00
HB=f(PC1,PC2)	0,59	21,01	0,02	<0,00

The ability to predict cucumber pigments by reducing the resulting feature vector with kPCA with a Gaussian kernel was tested. After removing the insignificant coefficients from the main models that have p-value >> α, it was found that the relationship between the dependent variables – pigments and the independent variables, can be described by the following models:

$C=f(PC_1,PC_2)$	$C = 0{,}41 + 0{,}28{PC_1}^2 - 0{,}55PC_1PC_2$	(49)
$X=f(PC_1,PC_2)$	$X = 0{,}56 - 0{,}14{PC_1}^2 - 6{,}1{PC_2}^2$	(50)
$HA=f(PC_1,PC_2)$	$HA = 0{,}49 - 0{,}26{PC_1}^2 - 5{,}05{PC_2}^2$	(51)
$HB=f(PC_1,PC_2)$	$HA = 0{,}03 + 4{,}2PC_2 - 0{,}13{PC_1}^2 - 16{,}57{PC_2}^2$	(52)

Table 7 shows the values for estimating the parameters of the obtained models. It can be seen that the highest value of the coefficient of determination and the lowest value of the errors has the model for predicting xanthophyll and the value of the critical coefficient is Fcr (2,29) = 3,33.

Table 7. Evaluation of the obtained models (kPCA with a Gaussian kernel) for pigments of cucumbers with watering

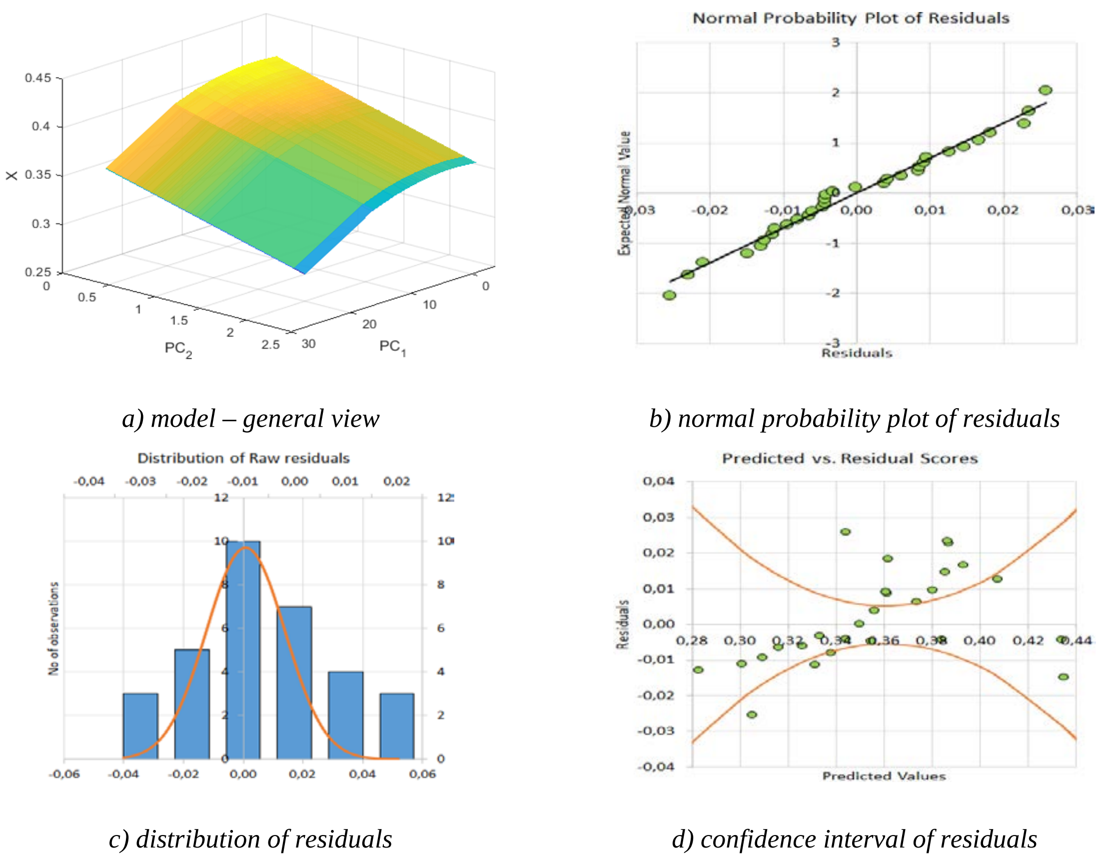
Model	R^2	F(2,29)	SE	p
C=f(PC1,PC2)	0,76	45,8	0,02	<0,00
X=f(PC1,PC2)	0,78	52,42	0,03	<0,00
HA=f(PC1,PC2)	0,69	32,64	0,03	<0,00
HB=f(PC1,PC2)	0,77	30,76	0,02	<0,00

The analysis of cucumber pigments during watering revealed that the xanthophyll pigment can be accurately predicted using a vector of features reduced by principal components with a polynomial kernel. The regression model is represented by equation (46).

Figure 8 displays the obtained model, where the values of the first principal component are plotted on the "x" axis, the values of the second principal component on the "y" axis, and the values of the xanthophyll pigment on the "z" axis. The nonlinearity of the relationship between these variables is apparent.

Residual analysis was also conducted. The residuals, being closely aligned to a straight line, indicate a distribution that approximates normality. Consequently, it can be assumed that the prerequisites for regression analysis are met.

Furthermore, Table 8 provides a summary of the results obtained for cucumbers with and without watering, regarding the determination of their pigments. Accuracy is measured by the coefficient of determination of the obtained model, while error represents the standard deviation of the respective model, reflecting the closeness to experimental data. Notably, higher accuracy in predicting pigments is achieved for cucumbers without watering. It is worth mentioning that the kPCA with linear and Gaussian kernels yield lower accuracy and relatively higher error rates in both groups of cucumbers with and without watering.



a) model – general view

b) normal probability plot of residuals

c) distribution of residuals

d) confidence interval of residuals

Figure 8. Residual analysis of a model of type X = f (PC1, PC2) obtained by kPCA-polynomial kernel

Table 8. Summary analysis of the obtained results

Type of cucumbers		Without watering		With watering	
Method for reducing the volume of data	Pigment	Accuracy, %	Error, %	Accuracy, %	Error, %
kPCA Simple	C	95%	2%	64%	2%
	X	95%	2%	89%	2%
	HA	94%	2%	65%	3%
	HB	94%	2%	67%	2%
kPCA Polynomial	C	96%	2%	72%	2%
	X	96%	2%	93%	1%
	HA	96%	2%	64%	3%
	HB	96%	1%	59%	2%
kPCA Gaussian	C	96%	2%	76%	2%
	X	96%	2%	78%	3%
	HA	96%	2%	69%	3%
	HB	95%	1%	77%	2%

C-carotene; X-xanthophyll; HA-chlorophyll A; HB-chlorophyll B

A comparative analysis of the obtained results with those from the available literature is made. Riccardi et al. [3] used colour digital images to obtain data that are correlated with chlorophyll content. The disadvantage of the method used by the authors is that it uses data from the RGB model, which is hardware dependent, and the values obtained are largely influenced by the conditions of obtaining digital images. This shortcoming is eliminated in the present work because colour indices from the Lab colour model, which is hardware-independent, are used. Spectral indices are also applied. The combined use of colour and spectral indices increases the accuracy to 96% and reduces the error (1-3%) in determining pigments in cucumbers by the non-contact method.

Liu et al. [5], offer a system for measuring chlorophyll. The system uses spectral characteristics in the range of 380-870 nm as input data. The post-processing of the spectral data is performed on a personal computer with the methods of partial least squares regression (PLSR), multiple linear regressions (MLR), and the method of support vector machines (SVM). The accuracy of chlorophyll determination is 67-97%. A disadvantage of this system is that the chlorophyll content cannot be determined directly in the field. Post-processing on a personal computer is required. Similar accuracy (59-96%) has been achieved in the present work. Relatively simplified computational procedures are used, which are suitable for use in microprocessor devices with low power consumption.

Friedrichs et al. [20], offer a device that works with fluorescent light, connecting to a smartphone, through which the content of chlorophyll A is measured. This device has achieved an accuracy of 98%. The disadvantage of this device is that it requires additional electronic equipment to be connected to the phone. Sample preparation is also required.

The tools proposed in this paper do not require connection equipment to a mobile phone.
This eliminates the disadvantage of this type of system.

Pineda-Tobón et al. [7], offer a solution for the indirect determination of the required amount of nitrogen by the chlorophyll content of plants. The authors offer a device that connects to a smartphone. It is achieved with an accuracy of 91%. The disadvantage of this device is that it requires additional electronic equipment to be connected to the phone. Sample preparation is also required. With the application in the present work, the combined use of colour and spectral indices and the reduction of the feature vector data with them, leads to a higher accuracy of up to 96%, without the need to connect an additional device to a mobile phone.

Vesali et al. [21] used a smartphone camera to determine the chlorophyll content in plants. In direct shooting, the authors get an accuracy of 59%. When the backlight is added, the accuracy increases to 97%. The authors use colour components from the RGB and HSV colour models and four colour indices separately. The disadvantage of the method used by them is that the color components and indices are used separately. The accuracy of the measurement is greatly influenced by the type of stage lighting. Higher accuracy would be achieved if combinations of different informative colour and spectral indices were used. This shortcoming has been eliminated in the present work by using a combination of features and presenting them in a new feature space. In this configuration of the recognition system, similar levels of accuracy of 59-96% are obtained.

From the comparative analysis made with technical solutions presented in the available literature, it can be considered that the results obtained in the present work complement and in some cases improve the solutions proposed by other authors.

Research using a mobile phone is mainly focused on the determination of chlorophyll in plants. In the present work, a detailed analysis is made of the possibility to predict not only chlorophyll but also the content of the pigments carotene and xanthophyll.

4. Conclusion

The obtained data sets containing 13 colour and 11 spectral indices are characterized by a large volume. This fact creates prerequisites for increasing the required processing time. This is not desirable. For this reason, informative colour features and spectral indices have been selected that are suitable for predicting plant pigment concentrations.

Tools have been developed for predicting cucumber pigments, which are based on the main colour and spectral indices, as well as on a certain set of relationships between them.

It has been found that the combined use of colour and spectral indices leads to an increase in the accuracy of predicting the values of cucumber pigments.

A total of 11 colour and 3 spectral indices can be used to predict pigments without of cucumbers watering. Also, 7 colour and 8 spectral indices can be used for cucumbers with watering. From these results, it was found that it is not possible to create average models for cucumbers grown in different conditions - with and without watering. This is because the colour and spectral indices that are suitable for predicting the characteristics of cucumbers are not the same under different conditions.

The obtained results complement and improve those presented in the available literature. Adapted and applied methods and tools require fewer computational procedures than the use of neural networks, which gives them the potential to be used in on-line monitoring of the condition of the cucumbers and their use in microprocessor systems with low power consumption.

Acknowledgements

This work was partially supported by the project, under contract №1.FTT/2021 "Development and research of a methodology for automated processing and analysis of data from electrical sensors using artificial intelligence techniques".

Also, the work was partially supported by the Bulgarian Ministry of Education and Science under the National Research Program "Intelligent plant growing" approved by DCM # 866/26.11.2020.

References:

[1]. Hashimoto, H., Uragami, C., & Cogdell, R. J. (2016). Carotenoids and photosynthesis. *Carotenoids in nature: Biosynthesis, regulation and function*, 111-139.

[2]. Dall'Osto, L., Fiore, A., Cazzaniga, S., Giuliano, G., & Bassi, R. (2007). Different Roles of α- and β-Branch Xanthophylls in Photosystem Assembly and Photoprotection. *Journal of biological chemistry*, *282*(48), 35056-35068.

[3]. Riccardi, M., Mele, G., Pulvento, C., Lavini, A., d'Andria, R., & Jacobsen, E. (2014). Non-destructive evaluation of chlorophyll content in quinoa and amaranth leaves by simple and multiple regression analysis of RGB image components. *Photosynthesis Research*, *120*, 263-272.

[4]. Madeira, A., Ferreira, A., de Varennes, A. & Vieira, M. (2007). SPAD meter versus tristimulus colorimeter to estimate chlorophyll content and leaf color in sweet pepper. *Communications in Soil Science and Plant Analysis*, *34*, 2461-2470.

[5]. Liu, B., Yue, Y.M., Li, R., Shen, W.J. & Wang K.L. (2014). Plant Leaf Chlorophyll Content Retrieval Based on a Field Imaging Spectroscopy System. *Sensors*, *14*(10), 19910-19925.

[6]. Sonobe, R., Yamashita, H., Mihara, H., Morita, A. & Ikka, T. (2020). Hyperspectral reflectance sensing for quantifying leaf chlorophyll content in wasabi leaves using spectral pre-processing techniques and machine learning algorithms. *International Journal of Remote Sensing*, *42*(4), 1311-1329.

[7]. Pineda-Tobón, D., Pérez, J., Gaviria-Palacio, D., & Guáqueta-Restrepo, J. (2017). Fast estimation of chlorophyll content on plant leaves using the light sensor of a smartphone. *Dyna*, *84*(203), 233-239.

[8]. Priyadarshini, R., Raj, G., & Shewade, D. (2016). Chromatography – the essence of bioanalysis. *European journal of biomedical and pharmaceutical sciences*, *3*(1), 366-377.

[9]. Baycheva, S. (2016). Application of devices of measurement of color in analysis of food products. *Innovation and entrepreneurship – Applied scientific journal*, *4*(4), 43-59.

[10]. Shalini, K. G., & Laxmi, A. (2007). Influence of additives on rheological characteristics of whole-wheat dough and quality of Chapatti (Indian unleavened Flat bread) Part I-hydrocolloids. *Food Hydrocolloids*, *21*(1), 110-117.

[11]. Pathare, P., Opara, U., & Al-Said, F. (2013). Colour measurement and analysis in fresh and processed foods: a review. *Food Bioprocess Technologies*, *6*, 36-60.

[12]. Vilaseca, M., Pujol, J., Arjona, M., & Martínez-Verdú, F. M. (2004, January). Color visualization system for near-infrared multispectral images. In *Conference on Colour in Graphics, Imaging, and Vision*, *2004*(1), 431-436. Society for Imaging Science and Technology.

[13]. Cermakova, I., Komarkova, J., & Sedlak, P. (2019). Calculation of Visible Spectral Indices from UAV-Based Data: Small Water Bodies Monitoring. *In 14th Iberian Conference on Information Systems and Technologies CISTI*, 1-5.

[14]. Atanassova, S., Nikolov, P., Valchev, N., Masheva, S., & Yorgov, D. (2019). Early detection of powdery mildew (Podosphaera xanthii) on cucumber leaves based on visible and near-infrared spectroscopy. *AIP Conference Proceedings*, *2075*, 160014-1-160014-5.

[15]. Taherdoost, H., Sahibuddin, S., & Jalaliyoon, N. (2022). Exploratory Factor Analysis; Concepts and Theory. *Advances in Applied and Pure Mathematics*, *27*, 375-382.

[16]. Kazlacheva, Z. (2019). Pattern design of twisted draperies with decorative and constructive function. *Applied Researches in Technics, Technologies and Education (ARTTE)*, *7*(1), 1-9.

[17]. Mladenov, M., Draganova, Ts., Dejanov, M., & Penchev, S. (2012). Complex Assessment of Grain Quality Fusing the Results from/of Color Image and Spectra Analyses. *International Federation of Automatic Control (IFAC)*, 127-132.

[18]. Čanađija, M., Brčić, M. & Brnić, J. (2013). Bending behavior of single-layered grapheme nanosheets with vacancy defects. *Engineering Review*, *33*(1), 9-14.

[19]. Chistokletov, N., & Vavilin, Y. (2018). Safety management system of machine-building production. *Engineering Review*, *38*(2), 226-231.

[20]. Friedrichs, A., Busch, J., Van der Woerd, H., & SmartFluo, O. (2017). A Method and Affordable Adapter to Measure Chlorophyll a Fluorescence with Smartphones. *Sensors*, *17*(4), 678.

[21]. Vesali, F., Omid, M., Mobli, H., & Kaleita, A. (2017). Feasibility of using smart phones to estimate chlorophyll content in corn plants. *Photosynthetica*, *55*(4), 603-610.

[22]. Daskalov, P., Kirilova, E., & Georgieva, T. (2018). Performance of an automatic inspection system for classification of Fusarium Moniliforme damaged corn seeds by image analysis. *MATEC Web of Conferences*, *210*, 02014

[23]. Alikhanov, J., Penchev, S., Georgieva, T., Moldazhanov, A., Kulmakhambetova, A., Shynybay, Z., Stefanov, E., & Daskalov, P. (2019). Design and performance of an automatic egg sorting system based on computer vision. *TEM Journal*, *8*(4), 319-1325.

[24]. Mladenov, M., Draganova, T., Tsenkova, R., & Mustafa, M. (2012). Quality assessment of grain samples using spectra analysis. *Biosystems Engineering*, *111*(3), 251-260.

[25]. Georgieva, T., Mihaylova, A., & Daskalov, P. (2020). Research of the possibilities for determination of some basic soil properties using image processing. In *7th International Conference on Energy Efficiency and Agricultural Engineering, 2020 - Proceedings*, 9279007.