

# Determination of optimal imaging parameters of hyperspectral images using the identification module

Andrey Chilikin<sup>1\*</sup> and Daniil Pupin<sup>1</sup>

<sup>1</sup>Federal Scientific Agroengineering Center VIM, build. 5, 1st Institute proezd, 109428, Moscow, Russia

**Abstract.** This paper is devoted to imaging with hyperspectral camera on the developed prototype of optical system for identification of micro- and macrodamage of plant tissues. The presented device is used for imaging process to detect fruit diseases at an early stage. Hyperspectral monitoring is a proven non-invasive method for detecting foetal damage. Hyperspectral imaging is a collection of spatially resolved spectral images representing an object or landscape. This data can be used to determine characteristics and parameters of fruit that indicate the presence of damage. This study reports on a triplicate experiment to determine the optimum parameters at which a hyperspectral image suitable for further detection and analysis of apples is obtained without the detection process slowing down the conveyor belt. This paper presents an optical identification device, which is a set of guides and racks on which a platform with a holder and a hyperspectral sensor attached to it moves. The recording and analysis of the hyperspectral cube was carried out using BaySpec's SpecGrabber and Gelion. Data processing of the planned experiment was carried out in the MatchCad 15 environment and the PlanExp B-D13 v.1.0 programme. To find the combination of factors giving the optimum value of the output parameter. In the course of calculations the results of statistical processing of the obtained data and the conclusion on the adequacy of the mathematical model were obtained, on the basis of which it was decided that the model is applicable for solving production problems and the extrema of the response function were obtained, which correspond to the values of optimal factors. The main parameters affecting the acquisition of hyperspectral images suitable for subsequent analysis are the speed of movement of the bracket with the camera, the illumination of the biological object and the height of the camera installation. These parameters are variable and their combination directly affects the image quality expressed in pixels.

## 1 Introduction

Sorting fruit for quality is the most important post-harvest handling operation. Fruits and vegetables are necessarily sorted before they are sent to storage or to the consumer. It is at

---

\* Corresponding author: [gwinnyandrei@yandex.ru](mailto:gwinnyandrei@yandex.ru)

this stage that it is possible to get rid of rejects, notice fruit samples that are not suitable in size or have damage due to phyto-diseases or mechanical impact. External damages include indentations, burns, bruises of fruits, which significantly reduce their marketability and seriously affect the final shelf life [1-4]. Unfortunately, sorting on the basis of mechanical and/or other damages is still often carried out completely manually - it is a monotonous and tedious work. Therefore, it can be said that finding alternative methods of fruit sorting is a demanded and urgent task. Having analysed the current state of affairs in the scientific sphere, it is possible to draw a conclusion about a number of problems hindering the development of technologies that contribute to the expansion of the application of automated control in the quality systems of fruit and vegetable products in the course of their sorting [5, 6]. First of all, it is connected with the lack of specific data on the quality of plant tissue of fruit and vegetable products on infoparameters of signals coming to sensors. Besides, it is extremely important task to ensure full coverage of quality control systems of the fruit surface, which got on the conveyor belt. One of the most important tasks is also the solution of the problem of qualitative determination of various damages received by fruits. Also, physical and biological variability of fruits cannot be excluded. Analysing the experience of the domestic industry, as well as scientific and technical works of foreign colleagues, one can conclude about the preference for optical-electronic control, as well as vision systems working with imaging in the visible and near-infrared spectral regions. Systems equipped with hyperspectral and multispectral technology for product quality control are also frequently used [9-14].

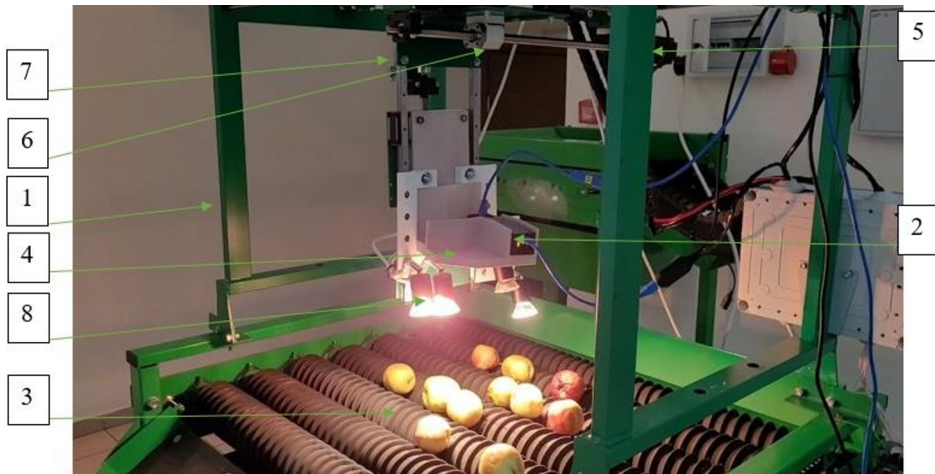
## **2 Materials and Methods:**

Hyperspectral reflectance images of apples were acquired using an original setup simulating a conveyor sorting device (Figure 1) equipped with a BaySpec OCI-F scanning hyperspectrometer (Bayspec Inc., San Jose, CA, USA). A reflectance spectrum (spectral range 400-1000 nm, spectral resolution 5-7 nm, 512 pixels per line) was determined for each pixel of the hyperspectral image and recorded using a reflectance standard from Spectralon. The camera was controlled remotely and the raw data were preprocessed using SpecGrabber and CubeCreator software. Four tungsten-halogen lamps (Camelion GU10 35W) were used as the illumination source for apple visualisation.

BaySpec SpecGrabber software is designed for hyperspectral imaging using BaySpec OCI™ cameras. It is intended to run on the Windows 7 operating system with 64 bits. BaySpec CubeCreator software is used to process original images from BaySpec hyperspectral cameras and convert them into hyperspectral cubes. It allows users to view and verify white calibration, dark background and original images, as well as display hyperspectral cubes and individual spectral bands. The software supports Windows 8, 7, Vista and XP operating systems. Gelion is software for analysing multi- and hyperspectral proximal sensing data. PlanExp B-D13 v.1.0 is a software product designed to calculate the planning matrix and process the output data of a three-factor experiment of B-D13 type. It is an important tool for solving scientific and applied problems related to the optimisation of properties of research objects, as well as the selection of formulation and technological parameters, where mathematical modelling using the method of orthogonal planning of experiments is applied [15-17].

The conveyor sorting device includes a hyperspectral camera, a control unit (laptop), a table with rubber rollers and a camera movement system. Hyperspectral camera 2 provides high detail of scanned objects. It emits light at various wavelengths in the infrared region of the spectrum and has spectroscopic characteristics such as reflectivity and fluorescence. Once the light is reflected from the surface of the object, the object is visualised. Depending on its chemical composition, the scanned object reflects light differently, creating a unique spectral

signature. This signature is then compared to reference data from a database stored on the control unit (laptop). The chamber 2 is fixed in a suspension 4, which allows it to move by means of stepper motors 5 along guides 6 and rack and pinion gears 7. It can move up and down in the vertical plane, as well as along and across in the horizontal plane, covering the entire working surface. Stepper motors 5 are powered by AC mains and are rigidly fixed to the frame of the sorting device 1, providing movement of the chamber through rack and pinion gears 7. The possibility of moving the camera 2 in horizontal and vertical directions allows full scanning of objects from different sides. The lighting system 8, located above the working surface, reduces the time required to scan objects. Due to the movable installation of hyperspectral camera 2 and lighting system 8 with microcontroller, the quality of detection of micro- and macro-damage of apple tissues is improved, as well as the accuracy of object classification during sorting.



1 - Chassis; 2 - hyperspectral camera; 3 - table with rubber rollers; 4 - camera suspension; 5,6 - stepper motors; 7 - transmission; 8 - light source.

**Fig. 1.** Installation for hyperspectral image of apple fruits.

In order to identify the most optimal parameters at which hyperspectral images suitable for processing and analysis were obtained, an experiment was carried out with a threefold repetition. The apple should be completely captured on the image and illuminated for further classification. Statistical processing of the obtained experimental results was carried out using mathematical modelling by the method of planning experiments. The main parameters that affect the quality of the hyperspectral image are: speed of the carriage with the camera; light flux coming from halogen lamps; height of the camera elevation above the sorting table. The selected parameters and intervals are presented in Table 1.

**Table 1.** Factors and their variation intervals.

	Lower level (-1)	Basic level (0)	Upper Level (+1)	Variation interval	Factor Name
x1:	0.02	0.04	0.06	0.02	Carriage speed. m/s
x2:	1300	1950	2600	650	Luminous flux. lm
x3:	0.2	0.3	0.4	0.1	Chamber lifting height. m

In the calculation of the factor plan, the values of input factor levels are taken in coded form, with the main level (the centre of the plan) of each factor designated as «0», and the

lower and upper levels as «-1» and «+1», respectively (Table 2). Image quality, which is expressed in the number of pixels, is taken as an output parameter.

**Table 2.** Experimental design and output parameters of the experiments.

Experiment number	The planning matrix			Natural values of the variables			Output parameter		
	x1	x2	x2	speed	light	height	y1	y2	y3
1	-1	-1	-1	0.02	1300	0.2	0.71	0.7	0.73
2	+1	-1	-1	0.06	1300	0.2	0.1	0.09	0.12
3	-1	+1	-1	0.02	2600	0.2	0.76	0.74	0.7
4	-1	-1	+1	0.02	1300	0.4	0.82	0.81	0.84
5	-1	0.19	0.19	0.02	2073.5	0.319	0.87	0.86	0.83
6	0.19	-1	0.19	0.0438	1300	0.319	0.68	0.71	0.65
7	0.19	0.19	-1	0.0438	2073.5	0.2	0.9	0.89	1.1
8	-0.29	+1	+1	0.0342	2600	0.4	1	1.3	1.6
9	+1	-0.29	+1	0.06	1761.5	0.4	0.52	0.51	0.56
10	+1	+1	-0.29	0.06	2600	0.271	0.61	0.67	0.62

Recalculation of the given natural values of factors is made by linear interpolation of values:

$$x_i = X_i - \frac{x_{0i}}{\Delta X_i} \tag{1}$$

where  $x_i$  – value of i-factor in coded form,

$X_i$  – value of i-factor in natural form,

$\Delta X_i$  – variation interval of i-factor.

Then the coefficients of the mathematical model were calculated. After obtaining the mathematical model, the significance (differences from zero) of the model coefficients and its adequacy were checked. The coefficients were checked for significance using Student's criterion (t-criterion), which is calculated by the formula:

$$t_i = \frac{|b_i|}{S\{b_i\}} \tag{2}$$

where  $b_i$  – i-coefficient of the mathematical model,

$S\{b_i\}$  – standard deviation in the definition of coefficients.

The standard deviation in determining the coefficients of the response function was determined by the formula:

$$S\{b_i\} = \sqrt{C_i \cdot S_b^2} \tag{3}$$

where  $C_i$  – values given for plan B-D13 in Table 2,

$S_b^2$  – variance of reproducibility in parallel experiments (Table 3).

**Table 3.** Coefficients of the mathematical model equation.

b0	b1	b2	b3	b11	b12	b13	b22	b23	b33
0.998	-0.185	0.173	0.096	-0.36	0.121	-0.001	-0.086	0.042	0.083

The variance of reproducibility in parallel experiments was calculated by the formula:

$$S_b^2 = \frac{1}{N(m-1)} \sum_{u=1}^N \sum_{j=1}^m (y_{uf} - \bar{y}_u)^2, \tag{4}$$

where  $N$  – number of experiments in the plan, pcs,

$m$  – number of parallel measurements in each experiment, pcs,

$y_{uj}$  – value of the output parameter in the  $u$ -experiment,  $j$ -parallel measurement,

$\bar{y}_u$  – the average value of the output parameter in the  $u$ -experiment.

The calculated value of  $t$ -criterion is compared with the tabular  $t_{\text{tabl}}$  for the chosen level of significance (5%) and the given number of degrees of freedom  $N(m-1)$ . When  $t_1 < t_{\text{table}}$  the coefficient  $b_i$  is considered significant.

The adequacy of the mathematical model was checked by Fisher's criterion (F - criterion). For this purpose, the adequacy variance was calculated by the formula:

$$S_{\text{ad}}^2 = \frac{m}{N-n_s} \sum_{u=1}^N (\bar{y}_u - \widehat{y}_u)^2, \tag{5}$$

where  $n_s$  – number of significant coefficients, pcs,

$\widehat{y}_u$  – response value predicted by the mathematical model equation.

Fisher's criterion is calculated as a ratio:

$$F = \frac{S_{\text{ad}}^2}{S_b^2}. \tag{6}$$

The calculated value of F-criterion is compared with the tabular  $t_{\text{tabl}}$  for the chosen significance level (5%) and numbers of degrees of freedom  $N(m-1)$  and  $N-n_s$ . When  $F < t_{\text{tabl}}$ , the equation of the mathematical model is considered adequate.

Data processing of the planned experiment was carried out in MatchCad 15 environment and PlanExp B-D13 v.1.0 programme. To find the combination of factors providing the optimal value of the output parameter, the response surfaces were plotted. When constructing one of the factors was left at a fixed level, and the other two were changed within the limits specified in Table 1.

The equation of the mathematical model:

$$y = (0,998) + (-0,185) \cdot x_1 + (0,173) \cdot x_2 + (0,096) \cdot x_3 + (-0,36) \cdot x_1^2 + (-0,086) \cdot x_2^2 + (0,083) \cdot x_3^2 + (0,121) \cdot x_1 \cdot x_2 + (-0,001) \cdot x_1 \cdot x_3 + (0,042) \cdot x_2 \cdot x_3$$

The main indicators obtained as a result of statistical processing of data are presented in Table 4.

**Table 4.** Results of statistical processing of the obtained data.

Indicator	Meaning
The variance of reproducibility in parallel experiments:	0.011
Tabulated value of Student's criterion:	2.09
The variance of adequacy of the mathematical model:	0.026
Number of degrees of freedom at significant coefficients:	4
Tabulated value of Fisher's criterion:	2.87
Calculated value of Fisher's criterion:	2.36

The result of testing the statistical significance of the regression equation parameters is presented in Table 5.

**Table 5.** Student's criteria and significance of model coefficients (1/0 - significant/unsignificant).

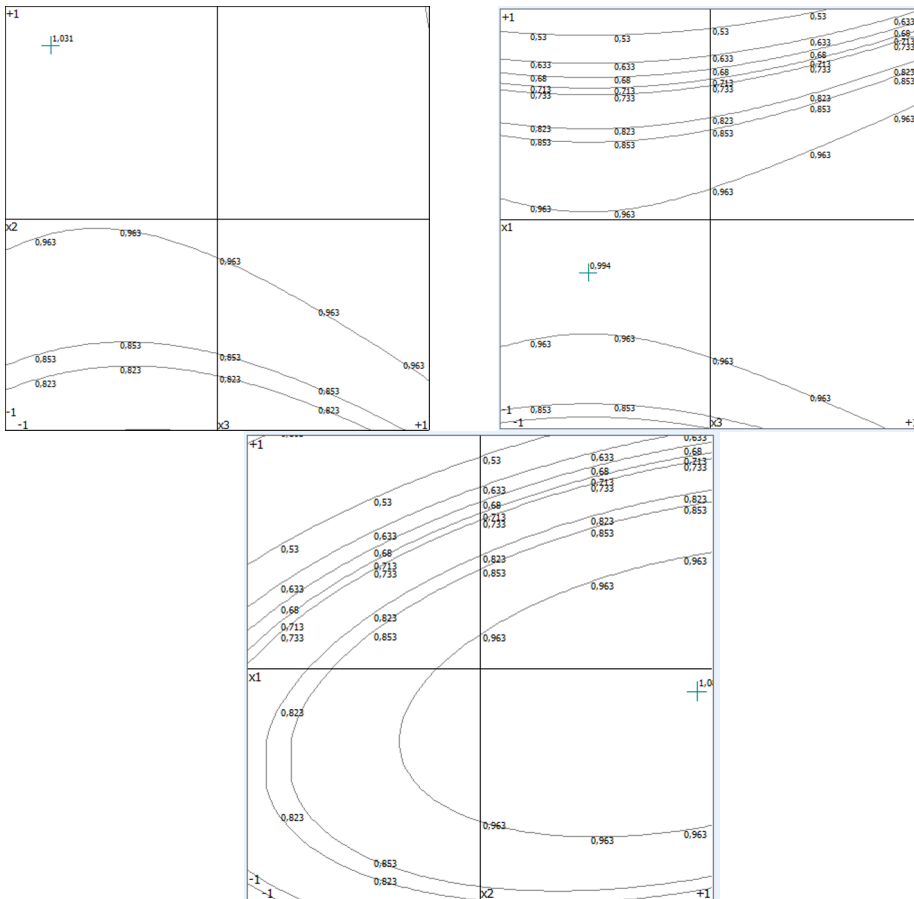
	b0	b1	b2	b3	b11	b12	b13	b22	b23	b33
t-criterion	10.273	4.449	4.161	2.309	4.479	2.441	0.02	1.07	0.847	1.033
Significance	1	1	1	1	1	1	0	0	0	0

### 3 Results and Discussion

Conclusion on the adequacy of the mathematical model: according to the Fisher criterion, the equation of the mathematical model is adequate. The model is applicable for solving production problems. As a constant factor is accepted:  $x_1=0$  (0,04). Transformed equation of the mathematical model taking into account the constant factor:

$$y = (0,998) + (0) + (0,173) \cdot x_2 + (0,096) \cdot x_3 + (0) + (-0,086) \cdot x_2^2 + (0,083) \cdot x_3^2 + (0) \cdot x_2 + (0) \cdot x_3 + (0,042) \cdot x_2 \cdot x_3$$

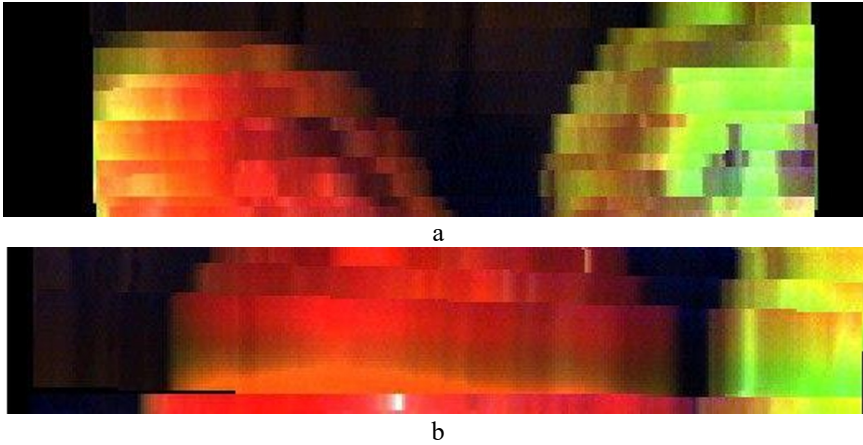
The equation of the mathematical model is a quadratic function of three variables. For the purpose of visual simplification and convenience of working with the mathematical model, a graphical interpretation of the function of three variables and a diagram of lines of equal level (isolines), which are projections of three-dimensional surfaces on the plane, have been constructed (Figure 2).



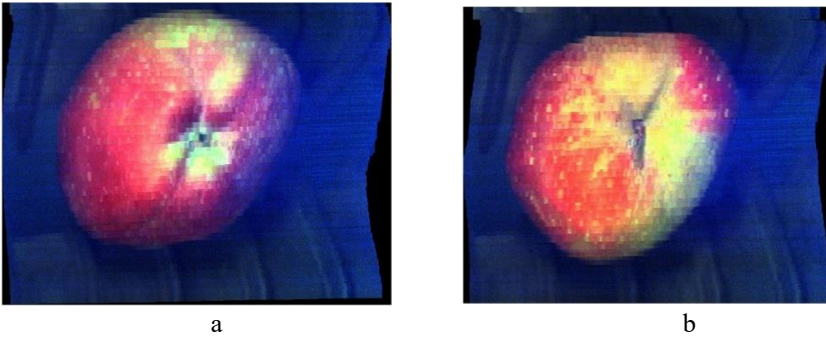
**Fig. 2.** Diagrams of equal level lines.

Presence of the extremum of the response function: the extremum of the response function is within the variation of the variables of the factors. The value of the extremum is  $Y_{opt}=1.031$ . The extremum of the response function corresponds to the values of the factors:  $x_2 = 0.814$  (2479.1) and  $x_3 = -0.784$  (0.222) with  $x_1 = 0$  (0.04).

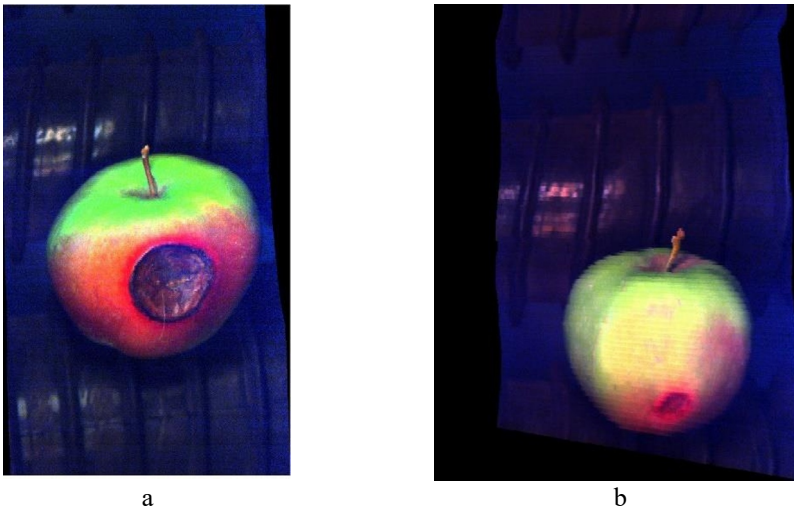
Figures 3, 4, 5 are shown as an illustrative example of the obtained result. At the above optimal parameters, images sufficient for processing and analysis are obtained.



**Fig. 3.** Result of imaging at high carriage speed low light and low camera position a-339 x 245 pixels, b - 401 x 259 pixels.



**Fig. 4.** Imaging result at optimum performance, a - 698 x 1099 pixels, b - 717 x 1060 pixels.



**Fig. 5.** Result of imaging at low carriage speed high light and high camera position a- 1853 x 1129 pixels, b- 2114 x 1260 pixels.

## 4 Conclusion

Based on the obtained data, we can conclude that the main parameters for obtaining hyperspectral images suitable for further analysis are: speed of movement of the arm with the camera, illumination of the biological object and height of the camera location. All parameters are variable and their combination directly affects the quality of images expressed in pixels. For this purpose, calculations were performed in a specialised software to obtain the optimal parameters. These parameters were used in practice and they proved to be sufficient for obtaining hyperspectral images suitable for analysis. Also, combinations were applied in practice, which allowed to obtain better quality images, but at the same time affecting the speed of the device, as they significantly reduce the speed of apples passing on the conveyor.

## References

1. A.V. Budagovsky et al, Application of laser irradiation for the estimation of functional state and the resistance to biotic and a biotic stresses in horticultural plants, EUCARPIA Fruit Breeding Section Newsletter, 31-32 (2001)
2. A.V. Budagovsky et al, Effect of spatial coherence flight on the photo regulation processes in cells, Physical Review E, 211-220 (2016)
3. A.V. Budagovsky et al, Cell response to quasi-monochromatic light with different coherence, Quantum electronics, 351-357 (2015)
4. A.V. Budagovsky et al, Application of laser irradiation for the estimation of functional state and the resistance to biotic and a biotic stresses in horticultural plants, EUCARPIA Fruit Breeding Section Newsletter, 31-32 (2001)
5. A.V. Budagovsky et al, Effect of spatial coherence flight on the photo regulation processes in cells, Physical Review E, 211-220 (2016)
6. A.V. Budagovsky et al, Cell response to quasi-monochromatic light with different coherence, Quantum electronics, 351-357 (2015)
7. Zhang Baohua, *State-of-the-art robotic grippers, grasping and control strategies, as well as their applications in agricultural robots* (Computers and Electronics in Agriculture, Moscow, 2020)
8. N.D. Egupov, *Methods of classical and modern theory of automatic control in five volumes* (Publishing House of Bauman Moscow State Technical University, Moscow, 2004)
9. S.L. Zenkevich, *Fundamentals of manipulative robot control. 2nd ed.* (Publishing House of Bauman Moscow State Technical University, M., 2004)
10. F. Mendoza, Grading of apples based on firmness and soluble solids content using Vis/SWNIR spectroscopy and spectral scattering techniques, Journal of Food Engineering, 59-68 (2014)
11. T. Ma et al, Tsuchikawa Rapid and nondestructive evaluation of soluble solids content (SSC) and firmness in apple using Vis-NIR spatially resolved spectroscopy, Postharvest Biology and Technology, 60-65 (2020)
12. C. Hill, *Learning Scientific Programming with Python* (Cambridge, UK, Cambridge University Press, 2020)
13. D. Drozdov et al, Supervisory Mutual Augmentation of Spectral Sensing and Machine Learning for Non-Invasive Detection of Apple Fruit Damages, Horticulturae, 2-13 (2023)

14. S.M. Vieira et al, *Cohen's Kappa Coefficient as a Performance Measure for Feature Selection. Proceedings of the International Conference on Fuzzy Systems, Congress on Computational Intelligence* (Barcelona, Spain, 2010)
15. F. Pedregosa et al, Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.*, 2825–2830 (2011)
16. J. Friedman et al, Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 2-5 (2010)
17. S. Walt et al, Scikit-image: Image processing in Python, *PeerJ*, 5-16 (2014)