Heliyon 9 (2023) e17389

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Rapid determination of triglyceride and glucose levels in *Drosophila melanogaster* induced by high-sugar or high-fat diets based on near-infrared spectroscopy

Jiamin Huang [a], Pengwei Wang [b], Yu Wu [a], Li Zeng [a], Xiaoliang Ji [b], Xu Zhang [a], Mingjiang Wu [a,**], Haibin Tong [a,***], Yue Yang [a,*]

[a] *Zhejiang Provincial Key Laboratory for Water Environment and Marine Biological Resources Protection, College of Life and Environmental Science, Wenzhou University, Wenzhou, 325035, China*

[b] *Key Laboratory of Watershed Science and Health of Zhejiang Province, School of Public Health and Management, Wenzhou Medical University, Wenzhou, 325035, China*

ARTICLE INFO

Keywords:
near-infrared spectroscopy
Drosophila melanogaster
Metabolic disorders
Triglyceride
Glucose
Rapid analysis

ABSTRACT

Triglyceride and glucose levels are important indicators for determining metabolic syndrome, one of the leading public-health burdens worldwide. *Drosophila melanogaster* is an ideal model for investigating metabolic diseases because it has 70% homology to human genes and its regulatory mechanism of energy metabolism homeostasis is highly similar to that of mammals. However, traditional analytical methods of triglyceride and glucose are time-consuming, laborious, and costly. In this study, a simple, practical, and reliable near-infrared (NIR) spectroscopic analysis method was developed for the rapid determination of glucose and triglyceride levels in an in vivo model of metabolic disorders using *Drosophila* induced by high-sugar or high-fat diets. The partial least squares (PLS) model was constructed and optimized using different spectral regions and spectral pretreatment methods. The overall results had satisfactory prediction performance. For *Drosophila* induced by high-sugar diets, the correlation coefficient (R_P) and root mean square error of prediction (*RMSEP*) were 0.919 and 0.228 mmoL gprot^{-1} for triglyceride and 0.913 and 0.143 mmoL gprot^{-1} for glucose respectively; for *Drosophila* induced by high-fat diets, the R_P and *RMSEP* were 0.871 and 0.097 mmoL gprot^{-1} for triglyceride and 0.853 and 0.154 mmoL gprot^{-1} for glucose, respectively. This study demonstrated the potential of using NIR spectroscopy combined with PLS in the determination of triglyceride and glucose levels in *Drosophila*, providing a rapid and effective method for monitoring metabolite levels during disease development and a possibility for evaluating metabolic diseases in humans in clinical practice.

1. Introduction

Metabolic syndrome refers to the co-occurrence of several metabolic risk factors including insulin resistance, obesity, atherogenic dyslipidemia, and hypertension. It is also a pathophysiological condition that includes type 2 diabetes, non-alcoholic fatty liver

* Corresponding author. College of Life and Environmental Science Wenzhou University Wenzhou, Wenzhou, 325035, China
** Corresponding author. College of Life and Environmental Science Wenzhou University Wenzhou, Wenzhou, 325035, China
*** Corresponding author. College of Life and Environmental Science Wenzhou University Wenzhou , Wenzhou, 325035, China
E-mail addresses: tonghb@wzu.edu.cn (M. Wu), wmj@wzu.edu.cn (H. Tong), yangyue2018@wzu.edu.cn (Y. Yang).

https://doi.org/10.1016/j.heliyon.2023.e17389
Received 17 March 2023; Received in revised form 12 June 2023; Accepted 15 June 2023
Available online 20 June 2023

disease, gout, cardiovascular disease, and chronic kidney disease [1,2]. Metabolic syndrome can cause elevated blood pressure, triglyceride, and fasting glucose, reduced high-density lipoprotein cholesterol, and central obesity. The disease will be diagnosed as metabolic syndrome if it has three or more symptoms mentioned above [3]. Almost 33.9% of the Chinese population and around a quarter of the world's population are estimated to suffer from the metabolic syndrome, which seriously affect the quality of human life [4–6].

A scientific and stable animal model is essential for the in-depth investigation of the pathogenesis of metabolic syndrome. Mice, zebrafish, and pigs are often used as animal models in studying metabolic syndrome, but they are not ideal due to long time consuming and high cost [7]. *Drosophila melanogaster* is a model organism that is suitable for studying human diseases due to its low cost, high reproducibility, simple genetic material, multiple mutant phenotypes, and easy observation [8,9]. In addition, *Drosophila* is becoming a popular animal model for investigating metabolic diseases because it has 70% homology to human genes and its regulatory mechanism of energy metabolism homeostasis is highly similar to that of mammals [10,11]. Currently, many studies have proved that high-sugar and high-fat diets can cause blood glucose rise, lipid disorders, obesity, and insulin resistance and eventually lead to type 2 diabetes and other chronic metabolic diseases in *Drosophila*. For example, *Drosophila* feeding a high-sucrose diet ingested less food and had decreased pupal rate, increased pupal mortality, and dramatically increased fat and glycogen levels [12]. Additionally, both high-sucrose and high-fat diet feeding can increase fat accumulation, lipid droplets, and the levels of glucose, trehalose, triglyceride and free fatty acid in *Drosophila* [13,14], confirming that high-sucrose or high-fat diet would impair glucose and lipid metabolism and cause *Drosophila* obesity, insulin resistance, and other metabolic diseases. Generally, the levels of triglyceride and glucose are the key detection indicators to make a preliminary determination of the health status of the organism, as well as a good predictor of diseases such as diabetes and metabolic syndrome. The common methods for analyzing triglyceride and glucose are enzymatic methods [15], thin-layer chromatography [16], coupled colorimetric assay [17], gas chromatography–mass spectrometry, and liquid chromatography–mass spectrometry [18]. Although they are accurate and stable, these methods have some disadvantages such as high cost, complex operation, long waiting time, use of organic reagents, and environmental pollution. Therefore, a faster, easier, and more efficient method to quantify the levels of triglyceride and glucose needs to be developed.

A good alternative is near-infrared (NIR) spectroscopy because of the following characteristics: the analysis speed is fast, and the spectrum collection of each sample can be completed in about 2 s; the NIR spectrum of the sample can be obtained by transmission or diffuse reflection, so the sample can be in any of three forms: solid, liquid and gas; for the NIR spectrum of a sample, multiple components in the sample can be analyzed at the same time; cost saving after purchasing the instrument for the first time and it has low technical requirements for operators; no reagent is needed during sample preparation, so it does not destroy the sample and pollute the environment [19]. NIR spectroscopy covers the spectrum region of 780–2526 nm, and the NIR absorption arises from the overtones and combinations of functional groups, such as *C*–H, *O*–H, and *N*–H bonds [20]. The NIR information can be associated with different chemical compositions because their molecules contain a large number of hydrogen-containing groups, which lays the theoretical basis for the qualitative and quantitative analysis of NIR spectroscopy [21]. Therefore, in NIR analysis, the development of a calibration model that correlates the NIR information and chemical parameters is the key step. The better the performance of the calibration model, the more accurate the prediction results are.

In the last decade, NIR spectroscopy has been successfully applied in agriculture [22], drug [23], foods [24,25], petroleum [26], textile [27], and environment [28]. And nowadays, NIR spectroscopy has been widely studied and applied for measuring biochemical parameters. Bazar et al. utilized NIR spectroscopy to predict total lipids, triglycerides, total cholesterol and high-density lipoprotein cholesterol in goose serum samples, and the correlation coefficients of prediction (R_P) were 0.864, 0.966, 0.850 and 0.793, respectively [29]. Filho et al. applied the NIR spectroscopy to determine the triglyceride level in human plasma. It was found that the multiple linear regression model exhibited the satisfactory performance, with the R_P and the root mean square error of prediction (*RMSEP*) were 0.986 and 20.00 mg dL^{-1}, respectively [30]. Neves et al. quantitatively analyzed the concentrations of glucose, triglyceride and high-density lipoprotein in rat plasma by NIR spectroscopy, and the *RMSEP* of PLS model were 6.08, 16.07 and 2.03 mg dL^{-1}, respectively [31]. These results showed that glucose, triglyceride and high-density lipoprotein in complex biological fluids can be simultaneously determined by NIR spectroscopy. Kasemsumran et al. applied NIR spectroscopy to determine glucose content in bovine serum samples. The R_P of the PLS model was 0.99, and the *RMSEP* was 25.31 mg dL^{-1} [32]. Xue et al. used on NIR spectroscopy and PLS model to measure the glucose content in mice, and the R_P and *RMSEP* were 96.22% and 0.419, respectively [33]. However, as for *Drosophila* model, only one study reported the determination of triglyceride levels and species through NIR spectroscopy [34]. And the experiment was conducted under the condition of normal growth of flies. To the extent of our knowledge, the quantification of triglyceride and glucose levels in *Drosophila* induced by high-sugar and high-fat diets via NIR spectroscopy has not been conducted.

This study aims to develop a rapid NIR methodology for evaluating triglyceride and glucose levels in *Drosophila* induced by high-sugar and high-fat diets. In the development of this methodology, the efficiency of different spectral regions and spectral pretreatment methods on the model performance was compared. This study provides an alternative method and potential clinical application for monitoring metabolite levels in individuals with metabolic diseases.

2. Materials and methods

2.1. Animals

W^{1118} *Drosophila melanogaster* was brought from ECIB-Fungene *Drosophila* Resource Center. Flies were reared in climatic chambers at a temperature of 25 °C ± 1 °C, humidity of 60% ± 5%, and 12/12 h light/dark cycle. Two types of diets were prepared for the experimental flies. The composition of the diets, including normal diet, high-sugar and high-fat diets, is listed in Table 1.

Table 1
Composition of the experimental diets.

Ingredients	Standard diets	High-sugar diets	High-fat diets
Water (mL)	100	100	100
Sucrose (g)	1.6	21.6	1.6
Glucose (g)	3.2	3.2	3.2
Corn (g)	8	8	8
Yeast (g)	3.2	3.2	3.2
Calcium chloride (g)	0.036	0.036	0.036
Ager (g)	0.7	0.7	0.7
Propionic Acid (g)	0.8	0.8	0.8
Palmitic acid (g)	0	0	2.5

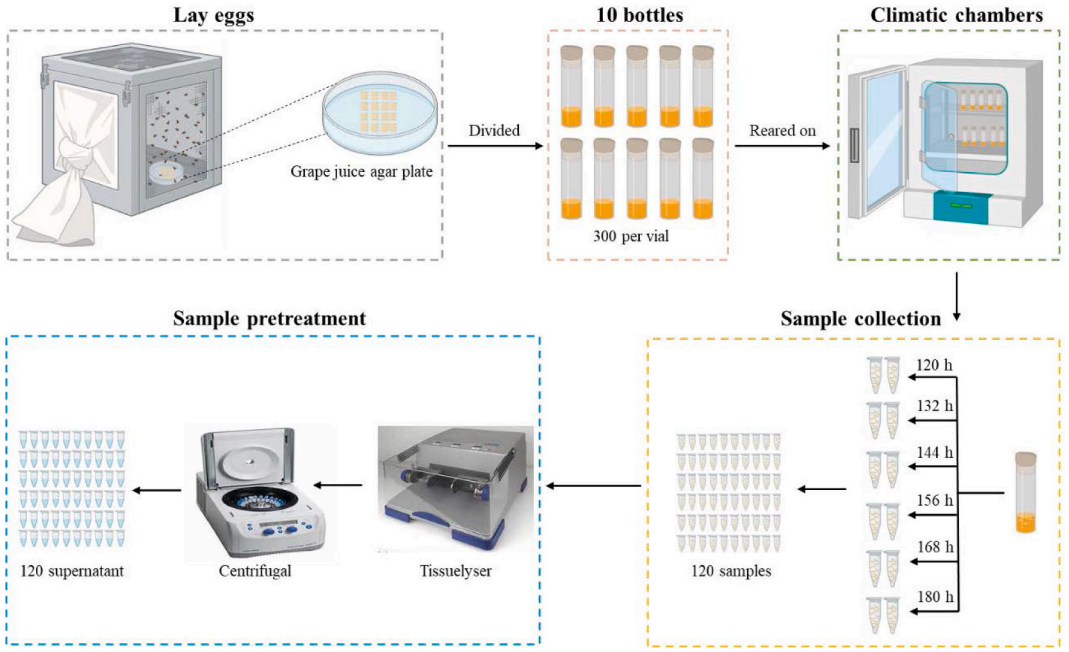

Fig. 1. Collection of *Drosophila melanogaster* samples induced by high-sugar diets.

2.2. Treatment and sample collection

Adult flies were permitted to lay eggs on grape juice agar plates with a modest number of yeast extract for 2 h at 25 °C. The collection process of *Drosophila* samples induced by high-sugar diets is illustrated in Fig. 1. Three thousand eggs were evenly divided into 10 bottles that were fed with high-sugar diets. Timing was started when the eggs were taken into the bottles. Different dietary conditions had different effects on the growth rate of *Drosophila* (Fig. 2). The high-sugar induced flies were collected at the 120th, 132nd, 144th, 156th, 168th, and 180th h. At each time point, 20 larvae were selected from a culture bottle and evenly distributed among two centrifuge tubes. Each centrifuge tube containing 10 larvae was considered as one sample. That is, 20 samples could be obtained at each time point. Finally, a total of 120 high-sugar induced *Drosophila* samples were collected. As for the high-fat induced *Drosophila*, the samples were collected at the 96th, 108th, 120th, 132nd, 144th, and 156th h. The other collection steps were the same as those in the high-fat samples. A total of 120 high-fat induced *Drosophila* samples were finally collected.

Each sample was washed with phosphate-buffered saline (PBS) to remove food residue on the surface and dried. Then, they were mixed with PBS +0.05% Tween-20. The mixture was centrifuged at 12,000 r/min for 15 min at a temperature of 4 °C. After centrifugation, the supernatant was stored in a 1.5 mL tube at −80 °C until analysis.

2.3. NIR spectral acquisition

The NIR spectra of *Drosophila* samples were collected using an Antaris II FT-NIR spectrometer (Thermo Fisher Scientific, USA) with RESULT software (version 8.0, Thermo Fisher Scientific, USA). The sample was placed into a quartz cell (1 mm optical path) using a disposable 1 mL syringe. After the NIR spectral acquisition, the quartz cell was cleaned three times with deionized water and dried with anhydrous ethanol. Each spectrum was recorded within the spectral region of 4000–10,000 cm^{-1} at a resolution of 8 cm^{-1}. The

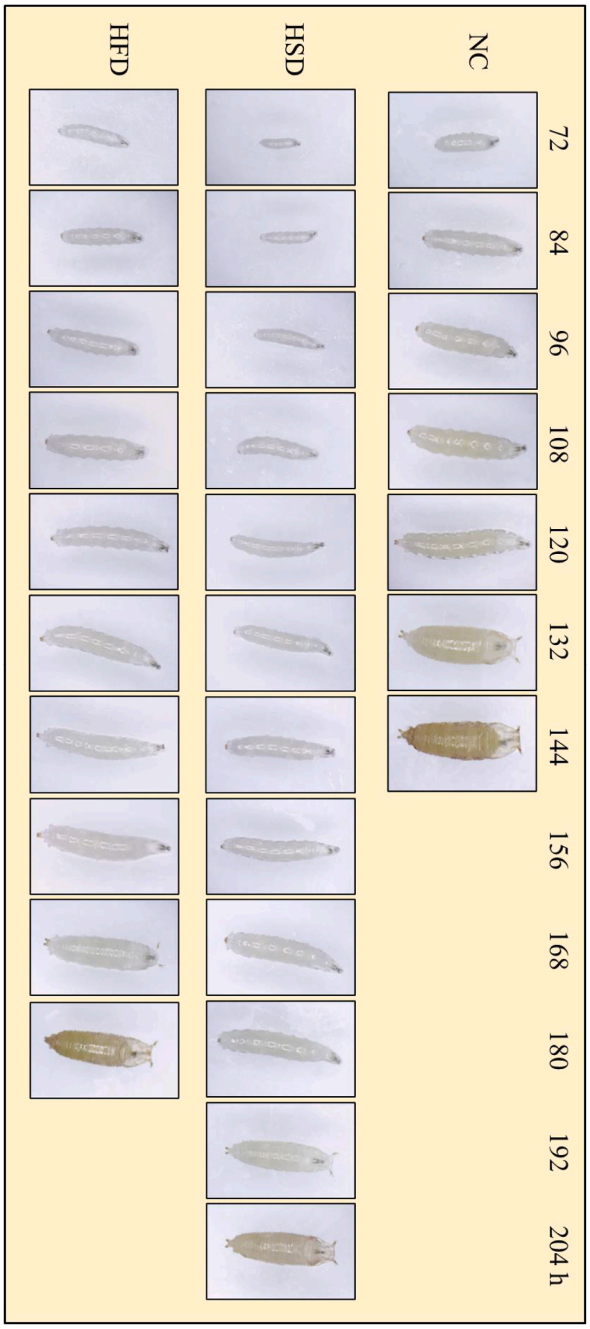

Fig. 2. The developmental morphology of *Drosophila melanogaster*. NC = normal control, HSD = high-sugar diets, HFD = high-fat diets.

absorbance spectra were obtained in the transmissive mode with air as reference standard and were acquired at room temperature (25 °C ± 1 °C). The measurement time for each spectrum was 60 s, and the background spectra were collected every 4 h. Each sample was scanned in triplicate, and the average spectrum was used for subsequent data analysis.

2.4. Determination of glucose and triglyceride

The determination of glucose and triglyceride levels was performed on an iMagic-V7 automatic biochemical analyzer (Shenzhen Icubio Biomedical Technology Co., Ltd., China). The obtained supernatant was taken into two tubes. Based on the biochemical analyzer, one tube was used to measure total protein and glucose using the corresponding commercial kits; the other tube was incubated for 5 min at 70 °C and then used to analyze the triglyceride using the triglyceride assay kit.

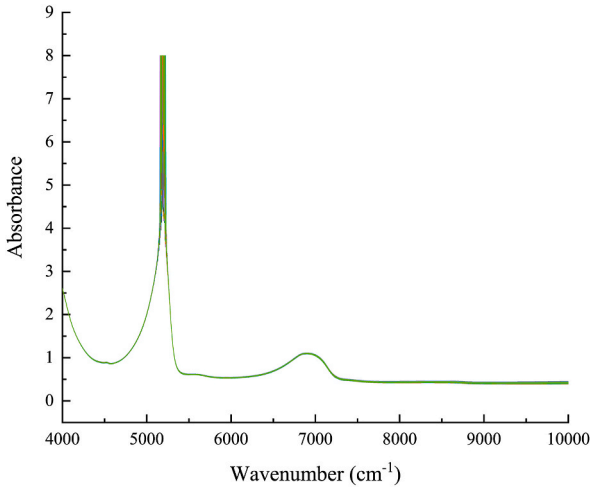

Fig. 3. Raw near-infrared spectra of 240 abnormal *Drosophila melanogaster* samples.

2.5. Data analysis

Data analysis was performed on the TQ Analyst software (Thermo Fisher Scientific, USA). The Mahalanobis distance was used to identify outliers. The concentration gradient method was applied for the partition of calibration and prediction sets. The PLS algorithm was used for the development of quantitative models of glucose and triglyceride. PLS can extract useful information as much as possible by simultaneously decomposing the independent variables (X) and dependent variables (Y) [35]. The PLS models were constructed based on the different spectral regions (4500–5500, 5500–6500, 6500–7500, 7500–10,000, and 4000–10,000 cm^{-1}), and their performances were compared. Multiplicative scatter correction (MSC), first derivative-Savitzky–Golay (1D/SG), first derivative-Norris (1D/Norris), second derivative-Savitzky–Golay (2D/SG), and second derivative-Norris (2D/Norris) were used as the spectral pretreatment methods. Additionally, the optimal number of factors that have a significant influence on the PLS model was determined using the cross-validation procedure.

The performance of quantitative models was evaluated using the root mean square errors of calibration (*RMSEC*) and prediction (*RMSEP*) and the correlation coefficient of calibration (R_C) and prediction (R_P) [36]. In general, the model performance is considered to be better if the R_C/R_P is closer to 1 and the *RMSEC/RMSEP* is closer to 0.

These indexes were calculated using the following equations:

$$Rc = \sqrt{1 - \frac{\sum (Ci - \widehat{C}i)^2}{\sum (Ci - Cn)^2}} \tag{1}$$

$$Rp = \sqrt{1 - \frac{\sum (Ci - \widehat{C}i)^2}{\sum (Ci - Cm)^2}} \tag{2}$$

$$RMSEC = \sqrt{\frac{\sum (\widehat{C_i} - C_i)^2}{n}} \tag{3}$$

$$RMSEP = \sqrt{\frac{\sum (\widehat{C_i} - C_i)^2}{m}} \tag{4}$$

where C_i refers to the reference values; $\widehat{C_i}$ refers to the NIR-predicted values; C_n is the average of the reference values in the calibration set; C_m is the average of the reference values in the prediction set; n is the number of samples in the calibration set; m is the number of samples in the prediction set.

3. Results and discussion

3.1. NIR spectral features

The biochemical molecules containing *C*–H, *N*–H, and *O*–H groups have a corresponding absorption band in the NIR spectral region, which lay the basis for qualitative or quantitative analysis. Fig. 3 shows the raw NIR spectra of 240 abnormal *Drosophila* samples within the wavenumber range of 4000–10,000 cm^{-1}. No significant differences in spectral characteristics were observed among

Table 2
Statistical results of the reference values of glucose and triglyceride in the calibration and prediction sets.

Group (N = 120)	Parameter (mmol gprot-1)	Calibration sets (N = 96)			Prediction sets (N = 24)		
		Min	Max	Mean	Min	Max	Mean
High-sugar diets	TG/TP	0.948	3.304	2.144	1.022	3.228	2.133
	GLU/TP	0.461	2.283	1.224	0.610	1.917	1.219
High-fat diets	TG/TP	0.786	1.893	1.293	0.849	1.722	1.289
	GLU/TP	0.706	2.717	1.502	0.880	2.348	1.509

TG = triglyceride, GLU = glucose, TP = total protein, N = number of samples, Min = minimum, Max = maximum.

Table 3
Effects of different spectral regions on the PLS models.

Group	Index (mmol $gprot^{-1}$)	Parameter	4500–5500 (cm^{-1})	5500–6500 (cm^{-1})	6500–7500 (cm^{-1})	7500–10000 (cm^{-1})	4000–10000 (cm^{-1})
High-sugar diets	TG/TP	R_C	0.666	**0.944**	0.789	0.861	0.666
		R_P	0.634	**0.879**	0.779	0.905	0.635
		RMSEC	0.433	**0.191**	0.356	0.294	0.432
		RMSEP	0.447	**0.276**	0.363	0.246	0.447
	GLU/TP	R_C	0.665	**0.928**	0.757	0.879	0.668
		R_P	0.653	**0.913**	0.763	0.907	0.652
		RMSEC	0.275	**0.137**	0.240	0.176	0.274
		RMSEP	0.266	**0.143**	0.227	0.148	0.266
High-fat diets	TG/TP	R_C	0.108	0.890	0.846	**0.936**	0.108
		R_P	0.150	0.836	0.599	**0.871**	0.150
		RMSEC	0.203	0.093	0.109	**0.072**	0.203
		RMSEP	0.195	0.108	0.158	**0.097**	0.195
	GLU/TP	R_C	0.864	0.395	0.409	0.459	**0.867**
		R_P	0.860	0.390	0.432	0.490	**0.853**
		RMSEC	0.157	0.286	0.284	0.277	**0.155**
		RMSEP	0.151	0.272	0.267	0.258	**0.154**

PLS = partial least squares, TG = triglyceride, GLU = glucose, TP = total protein, R_C = correlation coefficient of calibration, *Rp* = correlation coefficient of prediction, *RMSEC* = root mean square error of calibration, *RMSEP* = root mean square error of prediction.

various samples. Obviously, two strong absorption peaks were observed near 6944 and 5155 cm^{-1}, which arise from the first overtone of *O*–H and the combination of *O*–H stretching and deformation, respectively [37].

3.2. Outlier detection and sample partition

The outlier samples caused by the variations of measuring environment, instrument, and operation parameters can degrade model performance. The Mahalanobis distance, which can detect spectral errors, was used to identify outliers. No outlier sample was detected. Thus, all of the 120 high-sugar induced *Drosophila* samples and 120 high-fat induced *Drosophila* samples could be used for the subsequent model establishment and verification.

Sample division is essential to achieve a high-performance model. The samples should be divided into two subsets: the calibration set and the prediction set. The former is used to establish the model, whereas the latter is used to test the accuracy of the established model. To ensure that the division of the calibration set and prediction set was well proportioned, the content grads method was adopted [38]. The steps were conducted as follows: first, all the samples were sorted in accordance with the concentrations of the corresponding quality parameters; for every five samples, one sample was randomly selected into the prediction set, and the other four samples were assigned into the calibration set; finally, among the 120 *Drosophila* samples, 96 samples were used as calibration samples, and 24 samples were used as prediction samples. Table 2 summarizes the statistics of the reference values of glucose and triglyceride in the calibration and prediction sets. Regardless of the high-sugar or high-fat group, the ranges of the reference values of glucose and triglyceride in the calibration sets covered the ranges of those in the prediction sets, and the mean values of the two subsets were similar, indicating that the distribution of samples was uniform.

3.3. Selection of spectral region

The NIR spectra contain considerable redundant information that is irrelevant to the target attribute. Thus, the selection of characteristic spectral region is essential to obtain a model with good prediction accuracy [39]. As illustrated in Fig. 3, the NIR spectrum of *Drosophila* was a typical NIR spectrum of an aqueous solution, which had two strong absorption bands approximately occurring at the ranges of 4500–5500 and 6500–7500 cm^{-1}. These two absorption bands were mainly related to the *O*–H bond of water molecule, which arises from the combination of *O*–H stretching and deformation, and the first overtone of *O*–H, respectively [37]. Here, five different spectral regions, including 4500–5500, 5500–6500, 6500–7500, 7500–10,000, and 4000–10,000 cm^{-1}, were

Table 4
Effects of different spectral pretreatment methods on the PLS models.

Group	Index (mmol gprot^{-1})	Parameter	No	1D	1D/SG	1D/Norris	2D	2D/SG	2D/Norris
High-sugar diets	TG/TP	R_C	0.944	0.940	0.934	0.897	0.988	**0.992**	0.901
		R_P	0.879	0.898	0.894	0.837	0.853	**0.919**	0.850
		RMSEC	0.191	0.198	0.207	0.256	0.089	**0.072**	0.251
		RMSEP	0.276	0.255	0.259	0.317	0.302	**0.228**	0.305
	GLU/TP	R_C	**0.928**	0.856	0.855	0.928	0.852	0.839	0.899
		R_P	**0.913**	0.873	0.878	0.906	0.849	0.849	0.927
		RMSEC	**0.137**	0.190	0.191	0.137	0.192	0.200	0.161
		RMSEP	**0.143**	0.171	0.168	0.149	0.185	0.186	0.132
High-fat diets	TG/TP	R_C	**0.936**	0.957	0.980	0.927	0.778	0.803	0.960
		R_P	**0.871**	0.769	0.777	0.842	0.814	0.834	0.645
		RMSEC	**0.072**	0.059	0.041	0.077	0.128	0.122	0.057
		RMSEP	**0.097**	0.126	0.124	0.106	0.115	0.109	0.151
	GLU/TP	R_C	**0.867**	0.860	0.832	0.812	0.857	0.857	0.811
		R_P	**0.853**	0.865	0.856	0.832	0.830	0.850	0.833
		RMSEC	**0.155**	0.159	0.173	0.182	0.160	0.161	0.182
		RMSEP	**0.154**	0.148	0.153	0.164	0.165	0.156	0.164

TG = triglyceride, GLU = glucose, TP = total protein, R_C = correlation coefficient of calibration, *Rp* = correlation coefficient of prediction, *RMSEC* = root mean square error of calibration, *RMSEP* = root mean square error of prediction, 1D = the first derivative, 2D = the second derivative, SG = Savitzky Golay.

applied to construct the PLS models. According to Eqs. (1)–(4), the RC, RP, RMSEC, and RMSEP were obtained to evaluate the prediction performance of PLS models (Table 3). For the high-sugar group, the PLS models for triglyceride and glucose exhibited the best performance at the spectral region of 5500–6500 cm-1, both with R_P values greater than 0.87 and R_C values greater than 0.92. As for the high-fat group, the PLS model yielded the best performance in predicting the triglyceride levels at the spectral region of 7500–10,000 cm^{-1}, with R_P and *RMSEP* of 0.871 and 0.097 mmoL gprot^{-1}, respectively. The PLS model constructed based on the spectral region of 4000–10,000 cm^{-1} achieved the best results for the prediction of glucose levels, with R_P and *RMSEP* of 0.853 and 0.154 mmoL gprot^{-1}, respectively. Triglyceride is one of the component of lipids, with significant characteristic absorption at 1498–1878 nm, 1111–1265 and 1666–1818 nm [40].

The best spectral region for measurement of the glucose was reported at the wavelength range of 1500–1850 nm [41,42]. All these results agreed with the findings of this study.

3.4. Selection of spectral pretreatment method

Several uninformative information caused by background interference, environmental effects, and instrumental noise can weaken the prediction performance of the calibration model. Therefore, the preprocessing method should be performed on the raw NIR spectra to reduce the effects of interface factors and obtain a satisfiable prediction model [43]. Herein, seven different spectral preprocessing methods including no pretreatment, first derivative (1D), second derivative (2D) [31], 1D/SG, 2D/SG, 1D/Norris, and 2D/Norris were used. Based on the optimal spectral region, PLS models with different preprocessing methods were constructed. According to Eqs. (1)–(4), the R_C, R_P, *RMSEC*, and *RMSEP* were obtained to examine the model performances (Table 4). For the high-sugar group, the PLS model for triglyceride obtained the best prediction results as the NIR spectra were processed by the 2D/SG method; the PLS model for glucose optimized by the raw NIR spectra provided better results than those by other preprocessing methods. For the high-fat group, the raw NIR spectra without preprocessing produced superior performance for both triglyceride and glucose. Therefore, the PLS models for triglyceride and glucose of the high-sugar group should be established on the 2D/SG pretreated spectra and raw spectra, respectively. The raw NIR spectra without processing should be applied to construct the PLS models for triglyceride and glucose of the high-fat group.

3.5. Development of PLS models

PLS regression models were developed to create relationships between the spectrum data matrix and the target variables. In PLS, the factors that account for the most variations of the raw NIR spectra information have significant effect on the model performance. Few PLS factors can cause poor fitting between the reference values and the predicted values; more PLS factors can cause the "overfitting" problem, meaning that the model can yield good predictions in the calibration set but unsatisfactory fitting in the prediction set [44,45]. Thus, cross-validation was applied to determine the number of PLS factors, and which corresponds to the minimum values of the root mean square error of cross validation was regarded as the optimum one [46]. Finally, for the high-sugar group, the optimal numbers were 9 and 10 for triglyceride and glucose, respectively; for the high-fat group, the optimal numbers were 8 and 10 for triglyceride and glucose, respectively.

PLS quantitative models for triglyceride and glucose were developed based on the optimal spectral regions and preprocessing methods. The reference values and NIR predicted values for triglyceride (A and C) and glucose (B and D) in the high-sugar and high-fat groups using PLS models are illustrated in Fig. 4. The closer the data points to the red dashed line (1:1 line), the higher the prediction

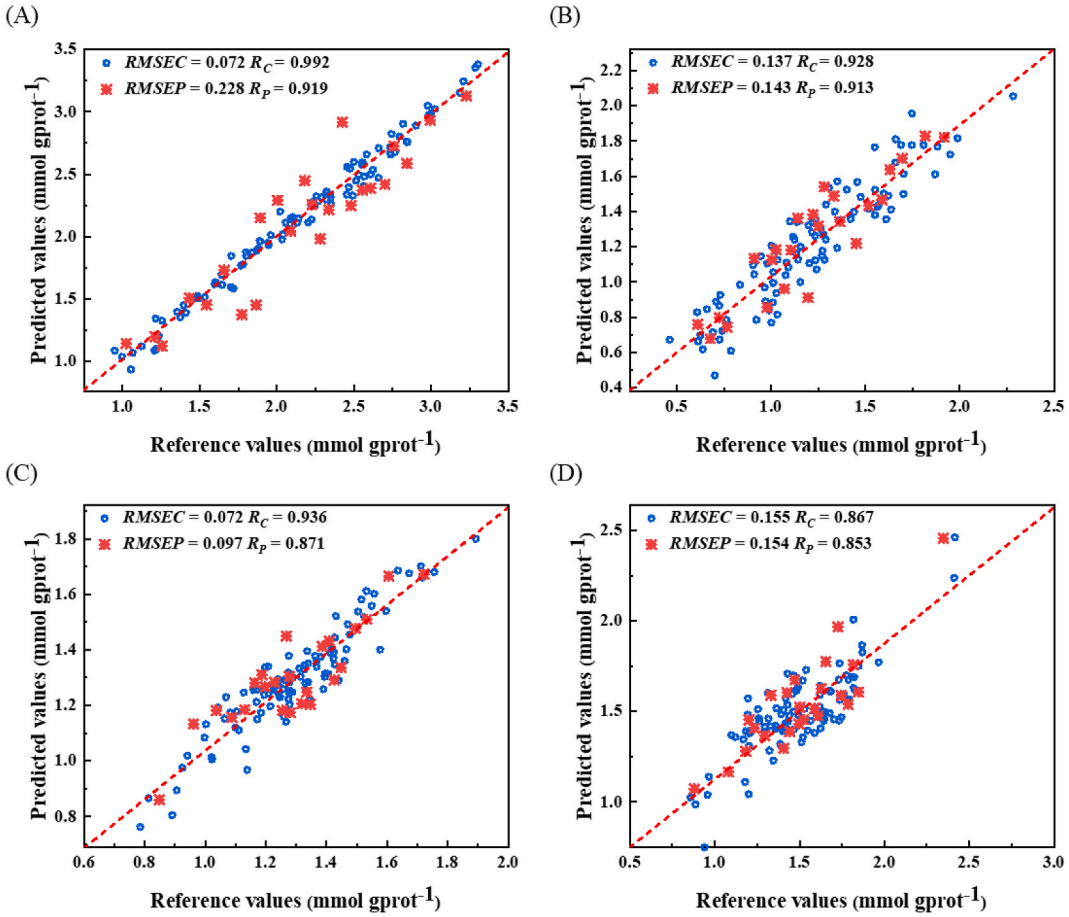

Fig. 4. Scatter plots of reference values and NIR predicted values for triglyceride (A and C) and glucose (B and D) in the high-sugar and high-fat groups using PLS models. The samples in calibration and prediction sets were distinguished by circles and asterisks, respectively. The red dashed lines represented the lines with slope = 1 was by. R_C = correlation coefficient of calibration, R_P = correlation coefficient of prediction, *RMSEC* = root mean square error of calibration, *RMSEP* = root mean square error of prediction. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

performance of the PLS model. All of the PLS models developed in this work were satisfactory. For the high-sugar group, the triglyceride and glucose could be well predicted by the PLS models. The R_P and *RMSEP* were 0.919 and 0.228 mmoL gprot^{-1} for triglyceride and 0.913 and 0.143 mmoL gprot^{-1} for glucose, respectively. For the high-fat group, the PLS models were relatively poor. The R_P and *RMSEP* were 0.871 and 0.097 mmoL gprot^{-1} for triglyceride and 0.853 and 0.154 mmoL gprot^{-1} for glucose, respectively. Fig. 5 presents the reference values and predicted values of triglyceride and glucose obtained from the PLS models during disease development. The PLS models were capable of capturing the changing trends for all indicators throughout the entire process. For the high-sugar group (Fig. 5A and B), the reference and predicted values were well superposed. For the high-fat group (Fig. 5C and D), some relative differences existed between the reference and predicted values of some individual sampling points, indicating that these PLS models should be further improved to achieve higher prediction performance. The good potential of NIR spectroscopy in the determination of biochemical parameters can also be found in some studies. For example, Han et al. demonstrated that human serum globulin could be satisfactorily predicted through NIR spectroscopy and optimal partner wavelength combination partial least squares (PLS) model, with a correlation coefficient of prediction (R_P) of 0.978 and root mean square error of cross validation of 0.813 g L^{-1} [47]. Yao et al. reported that NIR spectroscopy and equidistant combination PLS model can be used to analyze total cholesterol (TC) and triglyceride in human serum, with R_P and root mean square error of prediction (*RMSEP*) of 0.985 and 0.197 mmol L^{-1} for TC and 0.992 and 0.101 mmol L^{-1} for triglyceride, respectively [48].

4. Conclusions

This study developed an accurate and simple method that can rapidly measure the glucose and triglyceride levels in vivo of *Drosophila* induced by high-sugar or high-fat diets using NIR spectroscopy. PLS quantitative models based on different spectral regions and pretreatment methods were built, and their performances for predicting glucose and triglyceride were compared. For the

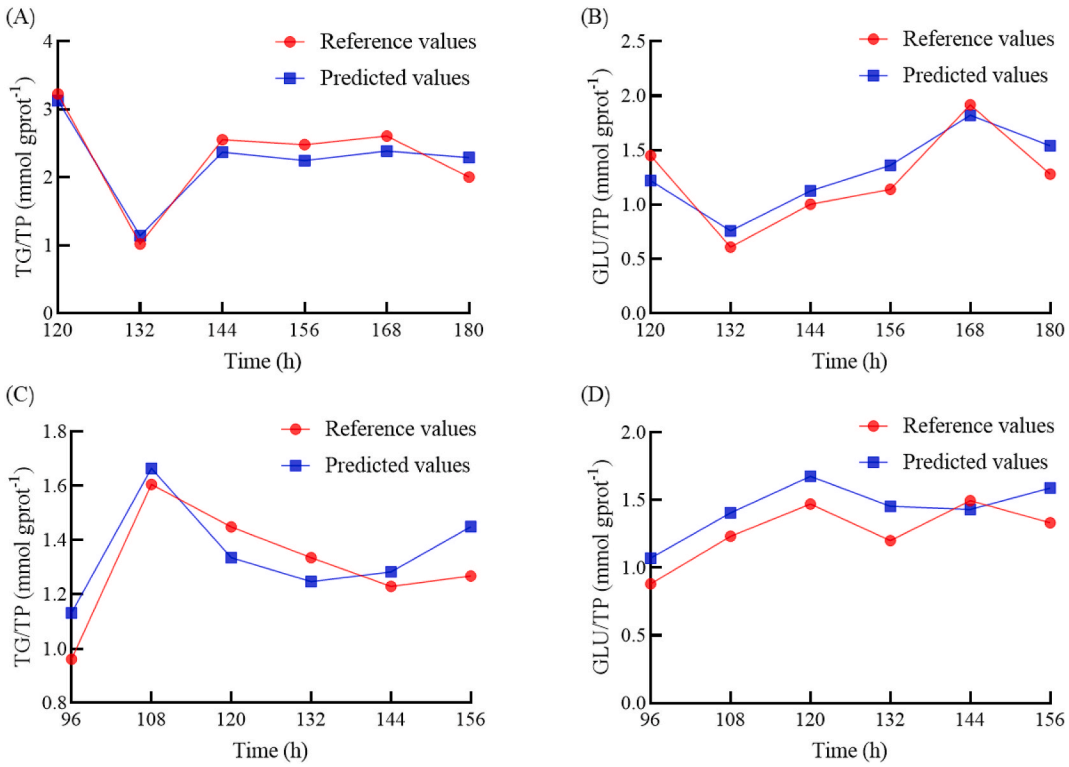

Fig. 5. Comparison of the NIR predicted values with the reference values for triglyceride (A and C) and glucose (B and D) in the high-sugar and high-fat groups.

triglyceride in the *Drosophila* induced by high-sugar or high-fat diets, the R_P values were both greater than 0.87; for the glucose, the R_P values were both greater than 0.85. The results confirmed the feasibility of NIR spectroscopy in the rapid determination of glucose and triglyceride levels. This method provides a rapid and effective way for monitoring metabolite levels during the development of metabolic diseases, which has huge potential for clinical application.

Author contribution statement

Jiamin Huang: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.
Pengwei Wang, Yu Wu, Xiaoliang Ji: Conceived and designed the experiments.
Li Zeng, Xu Zhang: Analyzed and interpreted the data.
Mingjiang Wu, Haibin Tong: Contributed reagents, materials, analysis tools or data.
Yue Yang: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

Yue Yang Yue Yang was supported by Zhejiang Provincial Natural Science Foundation of China {LTGN23C020001}, National Natural Science Foundation of China {31900274, 41876197, 81872952}.

Data availability statement

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R.H. Eckel, S.M. Grundy, P.Z. Zimmet, The metabolic syndrome, Lancet 365 (2005) 1415–1428, https://doi.org/10.1016/S0140-6736(05)66378-7.

[2] P.L. Huang, A comprehensive definition for metabolic syndrome, Dis. Model Mech. 2 (2009) 231–237, https://doi.org/10.1242/dmm.001180.
[3] K.G.M.M. Alberti, R.H. Eckel, S.M. Grundy, P.Z. Zimmet, J.I. Cleeman, K.A. Donato, J.-C. Fruchart, W.P.T. James, C.M. Loria, S.C. Smith, National heart, lung, and blood institute, American heart association, world heart federation, international atherosclerosis society, international association for the study of obesity, harmonizing the metabolic syndrome: a joint interim statement of the international diabetes federation task force on epidemiology and prevention; national heart, lung, and blood institute; American heart association; world heart federation; international atherosclerosis society; and international association for the study of obesity, Int. Diabetes Fed. Task Force Epidemiol. Prev. 120 (2009) 1640–1645, https://doi.org/10.1161/CIRCULATIONAHA.109.192644. Circulation.
[4] M. Aguilar, T. Bhuket, S. Torres, B. Liu, R.J. Wong, Prevalence of the metabolic syndrome in the United States, 2003-2012, JAMA 313 (2015) 1973, https://doi.org/10.1001/jama.2015.4260. –1974.
[5] Y.-J. Kwon, H.-S. Lee, J.-W. Lee, Association of carbohydrate and fat intake with metabolic syndrome, Clin. Nutr. 37 (2018) 746–751, https://doi.org/10.1016/j.clnu.2017.06.022.
[6] M.G. Saklayen, The global epidemic of the Metabolic syndrome, Curr. Hypertens. Rep. 20 (2018) 12, https://doi.org/10.1007/s11906-018-0812-z.
[7] T. Fuchs, M. de P. Loureiro, L.E. Macedo, D. Nocca, M. Nedelcu, T.A. Costa-Casagrande, Animal models in metabolic syndrome, Rev. Col. Bras. Cir. 45 (2018), e1975, https://doi.org/10.1590/0100-6991e-20181975.
[8] B. Ugur, K. Chen, H.J. Bellen, Drosophila tools and assays for the study of human diseases, Dis. Model Mech. 9 (2016) 235–244, https://doi.org/10.1242/dmm.023762.
[9] M. Yamaguchi, H. Yoshida, Drosophila as a Model organism, Adv. Exp. Med. Biol. 1076 (2018) 1–10, https://doi.org/10.1007/978-981-13-0529-0_1.
[10] E. Bier, R. Bodmer, Drosophila, an emerging model for cardiac disease, Gene 342 (2004) 1–11, https://doi.org/10.1016/j.gene.2004.07.018.
[11] P. Graham, L. Pick, Drosophila as a Model for diabetes and diseases of insulin resistance, Curr. Top. Dev. Biol. 121 (2017) 397–419, https://doi.org/10.1016/bs.ctdb.2016.07.011.
[12] B.M. Rovenko, O.I. Kubrak, D.V. Gospodaryov, N.V. Perkhulyn, I.S. Yurkevych, A. Sanz, O.V. Lushchak, V.I. Lushchak, High sucrose consumption promotes obesity whereas its low consumption induces oxidative stress in Drosophila melanogaster, J. Insect Physiol. 79 (2015) 42–54, https://doi.org/10.1016/j.jinsphys.2015.05.007.
[13] L.P. Musselman, J.L. Fink, K. Narzinski, P.V. Ramachandran, S.S. Hathiramani, R.L. Cagan, T.J. Baranski, A high-sugar diet produces obesity and insulin resistance in wild-type Drosophila, Dis. Model Mech. 4 (2011) 842–849, https://doi.org/10.1242/dmm.007948.
[14] N. Nayak, M. Mishra, High fat diet induced abnormalities in metabolism, growth, behavior, and circadian clock in Drosophila melanogaster, Life Sci. 281 (2021), 119758, https://doi.org/10.1016/j.lfs.2021.119758.
[15] C.M. Williams, R.H. Thomas, H.A. MacMillan, K.E. Marshall, B.J. Sinclair, Triacylglyceride measurement in small quantities of homogenised insect tissue: comparisons and caveats, J. Insect Physiol. 57 (2011) 1602–1613, https://doi.org/10.1016/j.jinsphys.2011.08.008.
[16] B. Al-Anzi, K. Zinn, Colorimetric measurement of triglycerides cannot provide an accurate measure of stored fat content in Drosophila, PLoS One 5 (2010), e12353, https://doi.org/10.1371/journal.pone.0012353.
[17] A. Hildebrandt, I. Bickmeyer, R.P. Kühnlein, Reliable Drosophila body fat quantification by a coupled colorimetric assay, PLoS One 6 (2011), e23796, https://doi.org/10.1371/journal.pone.0023796.
[18] J.M. Tennessen, W.E. Barry, J. Cox, C.S. Thummel, Methods for studying metabolism in Drosophila, Methods 68 (2014) 105–115, https://doi.org/10.1016/j.ymeth.2014.02.034.
[19] G.B. Rossi, V.A. Lozano, Simultaneous determination of quality parameters in yerba mate (Ilex paraguariensis) samples by application of near-infrared (NIR) spectroscopy and partial least squares (PLS), Lebensm. Wiss. Technol. 126 (2020), 109290, https://doi.org/10.1016/j.lwt.2020.109290.
[20] W.-Z. Lu, H.-F. Yuan, G.-T. Xu, Modern Near Infrared Spectroscopy Analytical Technology, China petrochemical press, Beijing, 2007.
[21] Y. Yang, R. Mao, L. Yang, J. Liu, S. Wu, M. Wu, X. Zhang, H. Tong, X. Ji, Rapid and comprehensive quality assessment of Bupleuri Radix through near-infrared spectroscopy combined with chemometrics, Infrared Phys. Technol. 121 (2022), 104051, https://doi.org/10.1016/j.infrared.2022.104051.
[22] E.S. de A. Duarte, V.E. de Almeida, G.B. da Costa, M.C.U. de Araújo, G. Véras, P.H.G.D. Diniz, D.D. de S. Fernandes, Feasibility study on quantification and authentication of the cassava starch content in wheat flour for bread-making using NIR spectroscopy and digital images, Food Chem. 368 (2022), 130843, https://doi.org/10.1016/j.foodchem.2021.130843.
[23] C.-C. Yang, P.-C. Yang, J.-J.J. Chen, Y.-H. Lai, C.-H. Hu, Y. Chang, S.J. Tu, L.-Y. Guo, Near-infrared spectroscopy for Monitoring sternocleidomastoid muscular oxygenation during isometric flexion for patients with mild nonspecific neck pain: a pilot study, Sensors 20 (2020) E2197, https://doi.org/10.3390/s20082197.
[24] Z. Guo, A.O. Barimah, L. Yin, Q. Chen, J. Shi, H.R. El-Seedi, X. Zou, Intelligent evaluation of taste constituents and polyphenols-to-amino acids ratio in matcha tea powder using near infrared spectroscopy, Food Chem. 353 (2021), 129372, https://doi.org/10.1016/j.foodchem.2021.129372.
[25] W. Tian, G. Chen, G. Zhang, D. Wang, M. Tilley, Y. Li, Rapid determination of total phenolic content of whole wheat flour using near-infrared spectroscopy and chemometrics, Food Chem. 344 (2021), 128633, https://doi.org/10.1016/j.foodchem.2020.128633.
[26] J.C.L. Alves, R.J. Poppi, Simultaneous determination of hydrocarbon renewable diesel, biodiesel and petroleum diesel contents in diesel fuel blends using near infrared (NIR) spectroscopy and chemometrics, Analyst 138 (2013) 6477–6487, https://doi.org/10.1039/c3an00883e.
[27] H. Chen, C. Tan, Z. Lin, Quantitative determination of wool in textile by near-infrared spectroscopy and multivariate models, Spectrochim. Acta Mol. Biomol. Spectrosc. 201 (2018) 229–235, https://doi.org/10.1016/j.saa.2018.05.010.
[28] R. Reda, T. Saffaj, S.E. Itqiq, I. Bouzida, O. Saidi, K. Yaakoubi, B. Lakssir, N. El Mernissi, E.M. El Hadrami, Predicting soil phosphorus and studying the effect of texture on the prediction accuracy using machine learning combined with near-infrared spectroscopy, Spectrochim. Acta Mol. Biomol. Spectrosc. 242 (2020), 118736, https://doi.org/10.1016/j.saa.2020.118736.
[29] G. Bazar, V. Eles, Z. Kovacs, R. Romvari, A. Szabo, Multicomponent blood lipid analysis by means of near infrared spectroscopy, Talanta 155 (2016) 202–211, https://doi.org/10.1016/j.talanta.2016.04.039, in geese.
[30] P.A. da C. Filho, R.J. Poppi, Determination of triglycerides in human plasma using near-infrared spectroscopy and multivariate calibration methods, Anal. Chim. Acta 446 (2001), https://doi.org/10.1016/S0003-2670(01)00956-4.
[31] A.C. de O. Neves, A.A. de Araújo, B.L. Silva, P. Valderrama, P.H. Março, K.M.G. de Lima, Near infrared spectroscopy and multivariate calibration for simultaneous determination of glucose, triglycerides and high-density lipoprotein in animal plasma, J. Pharm. Biomed. Anal. 66 (2012) 252–257, https://doi.org/10.1016/j.jpba.2012.03.023.
[32] S. Kasemsumran, Y.P. Du, K. Maruo, Y. Ozaki, Improvement of partial least squares models for in vitro and in vivo glucose quantifications by using near-infrared spectroscopy and searching combination moving window partial least squares, Chemometr. Intell. Lab. Syst. (2005) 82, https://doi.org/10.1016/j.chemolab.2005.08.014.
[33] J.T. Xue, L.M. Ye, Y.F. Liu, C.Y. Li, H. C, Noninvasive and fast measurement of blood glucose in vivo by near infrared (NIR) spectroscopy, Spectrochim. Acta Mol. Biomol. Spectrosc. 179 (2017) 250–254, https://doi.org/10.1016/j.saa.2017.02.032.
[34] W.C. Aw, J.W.O. Ballard, Near-infrared spectroscopy for metabolite quantification and species identification, Ecol. Evol. 9 (2019) 1336–1343, https://doi.org/10.1002/ece3.4847.
[35] G. Lee, K. Lee, Feature selection using distributions of orthogonal PLS regression vectors in spectral data, BioData Min. 14 (2021) 7, https://doi.org/10.1186/s13040-021-00240-3.
[36] H. Zhan, J. Fang, L. Tang, H. Yang, H. Li, Z. Wang, B. Yang, H. Wu, M. Fu, Application of near-infrared spectroscopy for the rapid quality assessment of Radix Paeoniae Rubra, Spectrochim. Acta Mol. Biomol. Spectrosc. 183 (2017) 75–83, https://doi.org/10.1016/j.saa.2017.04.034.
[37] H. Mark, J. Workman Jr., Chemometrics in Spectroscopy, Elsevier, 2010.
[38] Y. Yang, X. Liu, W. Li, Y. Jin, Y. Wu, J. Zheng, W. Zhang, Y. Chen, Rapid measurement of epimedin A, epimedin B, epimedin C, icariin, and moisture in Herba Epimedii using near infrared spectroscopy, Spectrochim. Acta Mol. Biomol. Spectrosc. 171 (2017) 351–360, https://doi.org/10.1016/j.saa.2016.08.033.
[39] Y. Wang, M. Li, L. Li, J. Ning, Z. Zhang, Green analytical assay for the quality assessment of tea by using pocket-sized NIR spectrometer, Food Chem. 345 (2021), 128816, https://doi.org/10.1016/j.foodchem.2020.128816.
[40] K.B. Beć, J. Grabska, C.W. Huck, Near-infrared spectroscopy in bio-applications, Molecules 25 (2020) E2948, https://doi.org/10.3390/molecules25122948.

[41] J. Liu, R. Liu, K. Xu, Accuracy of noninvasive glucose sensing based on near-infrared spectroscopy, Appl. Spectrosc. 69 (2015) 1313–1318, https://doi.org/10.1366/14-07728.
[42] L. Tang, S.J. Chang, C.-J. Chen, J.-T. Liu, Non-invasive blood glucose Monitoring Technology: a review, Sensors 20 (2020) E6925, https://doi.org/10.3390/s20236925.
[43] K. Heil, U. Schmidhalter, An evaluation of different NIR-spectral pre-treatments to derive the soil parameters C and N of a humus-clay-rich soil, Sensors 21 (2021) 1423, https://doi.org/10.3390/s21041423.
[44] F.C.B. Bedin, M.V. Faust, G.A. Guarneri, T.S. Assmann, C.B.B. Lafay, L.F. Soares, P.A.V. de Oliveira, L.M. Dos Santos-Tonial, NIR associated to PLS and SVM for fast and non-destructive determination of C, N, P, and K contents in poultry litter, Spectrochim. Acta Mol. Biomol. Spectrosc. 245 (2021), 118834, https://doi.org/10.1016/j.saa.2020.118834.
[45] J.-W. Hao, N.-D. Chen, C.-W. Chen, F.-C. Zhu, D.-L. Qiao, Y.-J. Zang, J. Dai, X.-W. Song, H. Chen, Rapid quantification of polysaccharide and the main onosaccharides in Dendrobium huoshanense by near-infrared attenuated total reflectance spectroscopy, J. Pharm. Biomed. Anal. 151 (2018) 331–338, https://doi.org/10.1016/j.jpba.2018.01.027.
[46] W. Wang, H. Dou, G. Zhang, L. Xie, Z. Wang, G. Deng, An approach for simultaneous monitoring the content of insensitive agent in the double-base oblate spherical propellant by application of near-infrared spectroscope and partial least squares, Spectrochim. Acta Mol. Biomol. Spectrosc. 258 (2021), 119851, https://doi.org/10.1016/j.saa.2021.119851.
[47] Y. Han, Y. Zhong, H. Zhou, X. Kuang, Optimal partner wavelength combination method applied to NIR spectroscopic analysis of human serum globulin, BMC Chem. 14 (2020) 37, https://doi.org/10.1186/s13065-020-00689-z.
[48] L. Yao, N. Lyu, J. Chen, T. Pan, J. Yu, Joint analyses model for total cholesterol and triglyceride in human serum with near-infrared spectroscopy, Spectrochim. Acta Mol. Biomol. Spectrosc. 159 (2016) 53–59, https://doi.org/10.1016/j.saa.2016.01.022.