INTRODUCTION

Cultivated mushrooms are in the focus of attention nowadays due to their economic and nutritional significance. Cultivated mushrooms are considered modern nutrient; the human body can get to significant functional nutritive value by consuming them. According to several researches, mushrooms are healthy sources of nutrient, have low calorie content, and are rich in proteins, fiber, vitamins, and minerals (Barroetaveña & Toledo, 2017; Urbain, Valverde, & Jakobsen, 2016).

The protein quality of mushrooms is better than most of the proteins from plant sources. The nutritive value of mushrooms is
primarily due to their protein content, a significant part of which is made up of essential amino acids. The amino acid composition of the proteins found in mushrooms is much more similar to animal proteins, thus they are more valuable than plant proteins (Meenu & Xu, 2019; Wang et al., 2014).

The carbohydrate content of mushrooms is important in a nutrition physiological respect, as it plays an important role in the composition of cell wall and energy supply. Recent studies have emphasized the importance of the carbohydrate content of mushrooms, as functional nutrition and bioactive components help to preserve health (Castro-Alves & do Nascimento, 2016; Rodrigues Barbosa, dos Santos Freitas, da Silva Martins, & de Carvalho, 2020).

Diffuse reflectance Fourier transform near infrared spectroscopy was employed for the identification of species and taxa delimitation of *Pleurotus* mushrooms. Principal component analysis (PCA) of FTIR spectra of *Agaricus bisporus* revealed that physical damage of mushrooms exhibits a significant effect on the tissue structure and aging process (Meenu & Xu, 2019).

The exact definition of the quantity of the nutrition-physiologically important components of mushrooms (protein, carbohydrate, and energy content) is a key task of quality control. Nowadays, classical wet chemical methods are used to define these components, which require a considerable amount of time, chemicals, and energy besides the significant impacts on the environment. (Table 1).

In modern analytical chemistry, examination methods based on chemometric and platform-free multivariate statistical processing of—"big data"—sets became more common. One of these is Fourier transform near-infrared spectroscopy (FT-NIRS), which makes the replacement of wet chemical methods possible being a fast, chemical-free, environmentally friendly, and green chemical method.

The purpose of this work was twofold. The first aim was to investigate and explain the differences between the mushroom species

| Method                      | Components | Value          | References                                      |
|-----------------------------|------------|----------------|-------------------------------------------------|
| *Agaricus bisporus*         | Dry weight | 14.3%          | Yildiz et al. (2017)                            |
| Traditional m. a            | Protein    | 3.2%           |                                                 |
| UV-VIS                      | Glucose    | 13.2%          |                                                 |
| NIR hyperspectral imaging   | Moisture   | RMSECV: 0.963% | Lin, Xu, & Sun (2019)                           |
| *Pleurotus ostreatus*       | Moisture   | 89.5–90.2      | Bhattacharjya et al. (2015), Yildiz et al. (2017) |
| Traditional m.              | Protein    | 25.3%–26.8%    |                                                 |
|                             | Carbohydrate | 39.7%–42.4%  |                                                 |
|                             | Fat        | 3.4%–4.5%      |                                                 |
| *Pleurotus eryngii*         | Protein    | 16.2%          | Rodrigues et al. (2015)                         |
|                             | Total carbohydrate | 64.9   |                                                 |
|                             | Fat        | 3.4%           |                                                 |
| *Lentinula edodes* b        | Moisture   | 91.5 ± 4.6%    | Li et al. (2018)                                |
| Traditional m.              | Protein    | 28.4 ± 0.7%    |                                                 |
|                             | Carbohydrate | 42.0 ± 1.1%  |                                                 |
|                             | Fat        | 2.1 ± 0.06%    |                                                 |
| *Hypsizygus tessellatus* b  | Protein    | 33.9%          | Chauhan, Prasad, Rathore, and Sharma. (2017)    |
| Traditional m.              | Carbohydrate | 50.1%    |                                                 |
|                             | Fat        | 2.8%           |                                                 |
| *Hericium erinaceus* c      | Proteins   | 16.8%–19.2%    | Rodrigues et al. (2015)                         |
| Traditional m.              | Total carbohydrate | 59.6%–61.2% |                                                 |
|                             | Fat        | 2.9%–3.2%      |                                                 |

aTraditional wet chemical methods.
bThe data refer to the fresh weight.
cThe data refer to the dried weight.

**TABLE 1** Summary of the examples for the determination of moisture, protein, carbohydrate, and fat content in cultivated mushrooms (The data refer to fresh weight except *Pleurotus ostreatus*)
studied by the infrared spectrum. Linear discrimination analysis was used to discriminate between species. The second aim was to develop a quick and simple quality control method for cultivated mushrooms.

The quality control methods used were FT-NIRS and Partial Least Square Regressions (PLSR) for the prediction of macro components (moisture, protein, total carbohydrate) and Interval Partial Least Square Regressions (iPLSR) for the prediction of the energy content of mushrooms from various species and breed.

2 | MATERIALS AND METHODS

2.1 | Samples

Altogether 192 samples from six different species were examined: cultivated white and brown champignon (Agaricus bisporus) (77 samples), cultivated oyster mushroom (Pleurotus x hybrid) (24 samples), shiitake (Lentinula edodes) (24 samples), white and brown shimeji (Hypsizygus tessellatus) (16 samples), king oyster mushroom (Pleurotus eryngii) (24 samples), and lion’s mane mushroom (Hericium erinaceus) (27 samples).

We obtained most of the samples from the Department of Vegetable and Mushroom Growing of SZIU and the rest of them from commercial sources. The preparation and measurement of each sample was carried out in the same way.

2.2 | Chemicals

All reagents (high purity analytical reagents) were ordered from Reanal (Budapest, Hungary). Ultra-pure water (18.2 MΩ cm) was purchased from a Milli-Q system from Merck-Millipore (Milford, MA, USA).

2.3 | Methods

2.3.1 | Reference methods

Dry matter content

The collected, freshly chopped mushroom samples were dried to constant weight in a programmable air circulating drying oven (Memmert, Schwabach, Germany). Results were expressed as % w/w. The estimation function was based on the spectra of fresh samples.

Protein content

The measurement of the protein content of the mushroom samples was carried out according to ISO standards (ISO, 2013) with Kjeldahl digestion method. Calculated nitrogen was multiplied by 4.38 (Ferreira, Morales, & Barros, 2017). Results were expressed as % w/w.

Carbohydrate content

The measurement of the carbohydrate content of the dried samples was carried out by the Luff-Schoorl method (Gafta, 2018). The method is applicable to all starch hydrolysis products. The reducing sugars in the sample (hydrolyzed if necessary) are heated to the boiling point under standardized conditions with a copper II solution, which is partially reduced to copper I. The excess copper II is subsequently determined iodometrically. As a preparative process long term, acidic hydrolysis was used (25 g/L HCl, 100°C, 3 hr) before the polysaccharides quantification. Results were expressed as % w/w.

Total carbohydrate content

The total carbohydrate content was calculated from the ash content to the following equations (Gezer et al., 2016): Total carbohydrates (g) = 100 − (g protein + g fat + g ash). The ash content was determined after a 12 hr long time cremation at 600°C (Labor MIM, Budapest, Hungary). This method makes it possible to determine the polysaccharide components that act as a fiber and cannot be considered as hydrolysates of starch. Results were expressed as % w/w.

Energy content

The calculation of energy content was carried out in accordance with Regulation 1169/2011/EU of the European Council (European Commission, 2011). According to the resolution the fat and fiber content of the sample is needed for the measurement of energy content. Mushrooms typically have a low value of fat content (0.1 – 0.5 g/100 g fresh weight), thus we did not measure it but used data found in the special literature concerning the different species (Ferreira et al., 2017).

In the case of fiber content, we considered that all the polysaccharides (glycogen, chitin, beta-glucan, hemicellulose, and pectin derivatives) making up the fiber content of mushrooms can be broken down to monomers by long-time acid hydrolysis.

As only some of the listed polysaccharides are utilizable, the energy content was calculated based on the amount of hydrolysable polysaccharides.

Energy was calculated according to the following equations:

\[
\text{Energy (kJ)} = 17 \times (g \text{ protein} + g \text{ carbohydrate}) + 37 \times (g \text{ lipid})
\]

wherein the carbohydrate value is the data obtained after 3 hr of hydrolysis.

Given that the hydrolysate does not represent the total carbohydrate content, this value is defined as “usable” energy. Results were expressed as kJ/100 g.

2.3.2 | Spectral measurements and pre-processing

The samples were subsequently measured on a Bruker MPA-Multipurpose FT-NIR analyzer (Bruker Optik GmbH, Ettlingen,
German) equipped with a quartz beam splitter, an integrated Rocksolid interferometer, a PbS detector, working in the 12,500–3,800 cm⁻¹ range combined with OPUS 7.2 (Bruker Optik GmbH, Ettlingen, German) software. Absorption spectra were collected in diffuse reflectance mode. The resolution was 8 cm⁻¹. The scanner speed was 10 kHz and each spectrum was the average spectrum of 32 subsequent scans. The internal background was measured using the integrated gold-coated surface of the integrating sphere. For spectral pre-processing standard normal variate (SNV), multiplicative scattering correction (MSC), first derivative (1st), standard normal variate combined with first derivative (SNV + 1st), straight-line subtraction combined with first derivative (SLS + 1st), and multiplicative scatter correction combined with first derivative (MSC + 1st) were applied.

Standard normal variate (SNV) is a pre-treatment used quite often in near infrared to remove the scatter. It is applied to every spectrum individually. The average and standard deviation of all the data points for that spectrum is calculated. Every data point of the spectra is subtracted from the mean and divided by the standard deviation (Bi et al., 2016; Cen & He, 2007). Straight-line subtraction (SLS) fits a straight line to the spectrum and subtracts it. This accounts for a tilt in the recorded spectrum. Multiplicative scattering correction (MSC) is a Math treatment to correct the scatter in the spectra. MSC attempts to estimate the coefficient which describes the scattering by regressing the spectrum to correct onto a reference spectra. (Chen, Song, Tang, Feng, & Lin, 2013; Wu et al., 2019).

The Savitzky–Golay first derivative (1st) is a polynomial derivative filter. The method used a smoothing of the spectra prior to calculating the derivative in order to decrease the detrimental effect on the signal-to-noise ratio than conventional finite-difference derivatives would have (Lee, Liong, & Jemain, 2018; Purwanto, Sari, & Budiastra, 2015).

2.4 | Statistics

Principal Component Analysis (PCA) was used for detecting the spectral outliers and linear discriminant analysis (LDA) was used to classify the six species of mushrooms. Partial Least Squares Regression (PLSR) was used for building the prediction models and interval Partial Least Squares Regression (iPLSR) was used for building the prediction model of energy.

2.4.1 | Principal component analysis

PCA is a frequently applied method for multivariate overview analysis of spectral data; it is a method of data reduction or data compression. Applied to a data matrix of samples by variables it constructs new variables, known as principal components (PCs). The new variables are linear combinations of the original ones, with the weight vectors that define the combinations being of unit length and orthogonal (at right angles) to each other. The first new variable captures as much as possible of the variability in all the original ones, having been constructed to have maximum variance amongst all such linear combinations. PCA is performed using the matrix with variety medians in rows and all descriptors in columns, to produce a map of all varieties, which resumes the main characteristics of the data. The PCA is performed on the variance–covariance matrix (Sampaio et al., 2017).

UNSCRAMBLER X 10.4 (Camo, Oslo, Norway) was used to determine spectral outliers. PCA analysis was carried out with a NIPALS algorithm and the method was validated by seven-segment random cross-validation. The F-residuals versus Leverage statistics represent three different kinds of outliers. The residual statistics on the ordinate axis describe the sample distance to model, whereas the leverage describes how well the sample is described by the model. The high residual variance may be due to non-important regions of a spectrum, for instance. Samples with high leverages have a stronger influence on the model than other samples; they may or may not be outliers, but they are influential. An influential outlier (high residual + high leverage) is the worst case; it can, however, easily be detected using an influence plot.

2.4.2 | Linear discriminant analysis

LDA is a very common technique for dimensionality reduction problems as a pre-processing step for machine learning and pattern classification applications to find a linear combination of features that characterizes or separates two or more classes of objects or events. The LDA technique is developed to transform the features into a lower dimensional space, which maximizes the ratio of the between—class variance to the within—class variance, thereby guaranteeing maximum class separability (Tharwat, Gaber, Ibrahim, & Hassanien, 2017).

LDA, as implemented in STATISTICA 12.0 (Tulsa, Oklahoma, USA), has different options to select the significant variables for model building, such as forward stepwise, backward stepwise, or all effects. All effects model building method and X-scrambling randomization test were used five times.

2.4.3 | Partial least square regression

The idea behind PLSR is to find a few linear combinations (components or factors) of the original x-values and to use only these linear combinations in the regression equation. In this way, irrelevant and unstable information is discarded and only the most relevant part of the x-variation is used for regression. In chemometrical applications, covariates in regression models are often correlated, causing a colinearity problem that can be solved by partial least squares (PLS) regression. In addition, high dimensionality in the space of covariates is also a problem with more parameters than cases, a phenomenon usually found in chemical spectral data that can also be solved by PLS regression (Huerta, Leiva, Liu, Rodríguez, & Villegas, 2019).
The spectral data and the concentration data are written in the form of matrices, where each row in the spectral data matrix represents a sample spectrum. The concentration data matrix contains the corresponding concentration values of the samples. The matrices will be broken down into their eigenvectors, which are called factors or principal components. A PLS regression algorithm is used to find the best correlation function between spectral and concentration data matrix (calibration).

Cross-validation or test validation is used to verify the PLSR results. Statistical qualifying factors of PLS are the follow: RMSEC (Root Mean Square Error of Calibration), RMSECV (Root Mean Square Error of Cross-Validation), RMSEP (Root Mean Square Error of Prediction), \( R^2 \) (percentage of variance accounted for by the calibration), \( Q^2 \) (percentage of variance accounted for by the validation), and Rank (number of PLS vectors). These parameters were calculated with the well-known formulas (Lee et al., 2018; Pirouz, 2012).

OPUS version 7.2 software was applied to perform PLS regression.

### Interval partial least squares regression

2.4.4 | Interval partial least squares regression

The interval partial least squares regression (iPLSR) method is a common choice for variable selection especially in the case of near-infrared spectra because spectral data are highly correlated and the usage of variable windows is a better option than examination of each variable individually. This process is very similar to the original PLS method, but here the spectra are divided into a number of intervals (equal length or manually made intervals). The number of intervals is optional. PLS models are made for each interval and the aim of the method is to choose those few intervals which have the smallest RMSECV. The use of these intervals gives better prediction instead of using the whole spectrum (Islam et al., 2018; Rácz, Vass, Héberger, & Fodor, 2015; Xiaobo, Jiewen, Povey, Holmes, & Hanpin, 2010).

UNSCRAMBLER X 10.4 software was applied to perform iPLS regression.

### RESULTS AND DISCUSSION

3 | RESULTS AND DISCUSSION

Comparing the average spectra (Figure 1) of the studied species it can be seen that the spectra mainly differ in height. This difference was due to the different particle sizes and reflections deriving from it. The first derivative curve is an excellent way of detecting fine characteristic differences. However, no characteristic difference was observed here.

3.1 | Principal Components Analysis

Prior to creating a model, PCA must be performed in every case in order to examine spectral outliers (Figure 2). The ellipse indicates a 95% confidence interval.

The residual statistics on the ordinate axis (F-residuals) describe the sample distance to model, whereas the Hotelling’s \( T^2 \) describes how well the sample is described by the model.

Samples—with three exceptions—are located in the low residual + low leverage area. This proves that the different species studied belong to the same set of samples. Three samples (all three...
samples were lion's mane mushroom—*Hericium erinaceus*) have a high residual variance (lying in the upper regions of the plot), which is poorly described by the model. The high residual variance may be due to non-important regions of a spectrum, for instance. Based on F-residuals versus Hotelling’s $T^2$ statistics these three samples were not considered as real outliers.

### 3.2 Linear discriminant analysis

In the first step, the average spectra of the samples from 12,500 to 3,800 cm$^{-1}$ were used for principal component analysis. No data pre-treatment was applied as data pre-processing. The first 20 PCA scores were used for further analysis with LDA.

Figure 3 shows the final result with the comparison of a typical example for the X-scrambling validation model. The six earlier mentioned groups can be clearly classified based on LDA applied to FT-NIR spectra. The validation of the model returned good results as well. The correct classification rate of the cross-validated model was 96.5%.

About 143 samples were used for the LDA. In this way, it was possible to ensure that the sample numbers of the six species were approximately equal. Otherwise, a large number of *Agaricus bisporus* samples would distort the result.

The confusion matrix (Table 2) demonstrates the hit probabilities. There is a misclassification of *Pleurotus ostreatus* and *Pleurotus eryngii*. This can be explained by the fact that they both belong to the *Pleurotus* genus within the Pleurotaceae family.
3.3 | Partial least squares regression

We defined the dry material, protein, carbohydrate, and total carbohydrate content of the samples according to the procedure described in 2.3.1. and then by this data, we calculated the energy content in the described way. Each parameter was defined by three parallel measurements.

The calibration models were made for the original dataset, which contains 192 samples and the spectrum range between 12,500 and 3,800 cm\(^{-1}\).

### TABLE 2 Confusion matrix of LDA

|   | AB  | PO  | PE  | LE  | HT  | HE  |
|---|-----|-----|-----|-----|-----|-----|
| AB | 25  | -   | -   | -   | -   | -   |
| PO | -   | 23  | 1   | -   | -   | -   |
| PE | -   | 4   | 23  | -   | -   | -   |
| LE | -   | -   | -   | 24  | -   | -   |
| HT | -   | -   | -   | -   | 16  | -   |
| HE | -   | -   | -   | -   | -   | 27  |

The best estimation functions are summarized in Table 3. The selection of the evaluation ranges was based on the characteristic vibration areas of the examined macro component (Workman & Weyer, 2012).

Cross-validation and test set validation methods were used to test the calibration model.

The \(Q^2\) value of cross-validations was above 0.9 for both dry matter, carbohydrate, total carbohydrate, and protein content. The \(Q^2_T\) value of test set validations are lower than these but except for the carbohydrate content, in all cases, we obtained more than 0.9 here too.

### TABLE 3 Performance parameters, scaling methods, number of latent variables (Rank), and evaluation ranges of the best models

| Component          | Concentration range, %w/w | Scaling methods | Calibration | Validation | Test validation |
|--------------------|---------------------------|----------------|-------------|------------|----------------|
|                    |                           |                | \(R^2\) | RMSEC, % w/w | Q^2 | RMSECV, % w/w | \(Q^2_T\) | RMSEP, % w/w | Rank | Evaluation range, cm\(^{-1}\) |
| Dry matter         | 5.5–12.4                  | MSC + 1st der  | 0.952 | 0.32       | 0.931 | 0.36       | 0.921 | 0.37       | 5    | 7,001–6,700 |
|                    |                           |                |           |            |       |            |       |            |      | 5,300–5,000 |
| Carbohydrate       | 5.4–36.3                  | SNV + 1st der  | 0.928 | 1.53       | 0.906 | 1.68       | 0.867 | 1.76       | 9    | 7,502–4,247 |
| Total carbohydrate | 54.2–78.1                 | 1st derivative | 0.948 | 1.39       | 0.937 | 1.50       | 0.936 | 1.50       | 5    | 11,100–6,897 |
|                    |                           |                |           |            |       |            |       |            |      | 6,303–3,800 |
| Protein            | 12.9–34.6                 | SLS + 1st der  | 0.997 | 0.25       | 0.987 | 0.47       | 0.974 | 0.81       | 7    | 6,900–4,397 |

Note: \(R^2\)—percentage of variance accounted for by the calibration; \(Q^2\), \(Q^2_T\)—percentage of variance accounted for by the fivefold cross and test set validation; RMSEC—root mean error of calibration; RMSECV—root mean square error of fivefold cross validation; RMSEP—root mean square error of test set validation; Rank—number of PLS vectors. It was chosen based on the global minimum of the root mean squared error of cross-validation (RMSECV).

### TABLE 4 Determination of utilized energy content—PLS-R data for the whole measurement range

| Component | Concentration range, kJ/100 g | Scaling methods | Calibration | Validation |
|-----------|--------------------------------|----------------|-------------|------------|
| Energy    | 545–1,128                      | No             | 0.787 | 55.7 | 0.698 | 66.7 | 10 |
| SNV       |                                |                | 0.829 | 49.9 | 0.706 | 66.1 | 10 |
| MSC       |                                |                | 0.873 | 43   | 0.772 | 58   | 10 |
| 1st       |                                |                | 0.926 | 32.9 | 0.519 | 84.2 | 6  |
| SNV + 1st |                                |                | 0.935 | 30.7 | 0.550 | 81.8 | 6  |
| MSC + 1st |                                |                | 0.903 | 37.5 | 0.518 | 84.3 | 5  |

Note: Best model was set in bold italics.
$Q^2$ of the prediction of carbohydrate content is the lowest value, which can be explained by the fact that the method of determining classical data (Luff Schoorl method) has a lot of uncertainties and errors.

As these results meet the general fit-for-purpose requirements for the quantification of macronutrients in cultivated mushrooms, the FT-NIR related methods can be adapted for routine applications.

Similar NIR prediction functions have been developed to determine the total polysaccharide content of some *Ganoderma* species, which can be characterized by RMSEP = 0.224% w/w and 0.603% w/w (Chen et al., 2012; Ma et al., 2018).

**FIGURE 4** (a) Variable selection with iPLS-R—20 segments. (b) Variable selection with iPLS-R—30 segments
Similar NIR prediction functions have been developed to determine the total polysaccharide content of some *Ganoderma* species, which can be characterized by RMSEP = 0.224% w/w and 0.603% w/w (Chen et al., 2012; Ma et al., 2018). Although their results cannot be compared with ours because the spectra of *Ganoderma* samples differ so significantly from the spectra of the examined cultivated mushroom species, that they were spectral outliers.

### 3.4 | Interval partial least squares regression

Given that energy content is a complex property—it can be calculated based on the common influence of several macro-components (carbohydrates, proteins, fat)—interval PLS regression (iPLSR) was used to predict the energy content.

As a first step, we completed the model building concerning the whole range of measurement (12,500–3,800 cm\(^{-1}\)) by different data pre-processing methods. In accordance with earlier models, a maximum of 10 latent variables was taken into consideration.

In this case, the checking of the models was also carried out by a sevenfold segment random cross-validation (Table 4).

We got the most favorable relations (best \(Q^2\) and least RMSECV) after the combined data pre-processing of SNV + first derivate.

During the iPLSR examination, the whole range of measurement was divided into 20 and 30 identical parts. The RMSECV value of each model (Figure 4a,b) was examined based on the PLS regression results of each area and the ranges with the lowest RMSECV values were selected.

The wavelength ranges were divided into 20 equal parts and the following ranges were selected: 8,867–8,135, 7,695–7,263, and 6,823–3,800 cm\(^{-1}\). By dividing the wavelength ranges into 30 equal parts, the following ranges were selected by method 30_1: 7,575–7,290, 6,129–5,554, and 5,261–3,800 cm\(^{-1}\).

We have examined the models that can be set up by extending the evaluation ranges and also considering two of those ranges where the RMSECV value is only slightly higher than the average RMSECV value. (Figure 4b). Using method 30_2, two areas that are only slightly above the average RMSECV were added to the 30_1 ranges: 8,733–8,447 and 8,443–8,158 cm\(^{-1}\).

The ranges above correspond from a chemical point of view to the typical vibration areas of components responsible for energy content (Workman & Weyer, 2012). PLS regression was repeatedly applied to the selected intervals.

**TABLE 5** Comparison of PLS and iPLS regression results for the utilizable energy content

|               | PLS | iPLS_20 | iPLS30_1 | iPLS30_2 |
|---------------|-----|---------|----------|----------|
| \(Q^2\)       | 0.549| 0.862   | 0.862    | 0.867    |
| RMSECV, kJ/100 g | 81.9 | 45.0    | 45.7     | 44.4     |
| Rank          | 5   | 8       | 8        | 8        |

**FIGURE 5** iPLS results for energy content
The results obtained by PLSR and iPLSR regression are summarized in Table 5.

By examining the ranges resulting from the iPLSR procedure it can be seen that in accordance with earlier presented models domains with large energy (12 500–9,000 cm\(^{-1}\)) do not contain useful information.

The most favorable result was achieved with the extended 30-units iPLS (30_2 method) (Figure 5).

Applied data preprocessing was SNV + first derivative combined procedure. To set up the PLS model, eight PLS components were needed.

The \( R^2 \) value was 0.907 for the calibration and the \( Q^2 \) value was 0.867 for the validation. The RMSEC value was 36.8 (kJ/100 g) for the calibration and RMSECV value was 44.4 (kJ/100 g) for the validation.

4 | CONCLUSION

We successfully applied the FT-NIRS technique combined with chemometrical methods to evaluate the quantitative characteristics of the nutritional values of cultivated mushrooms.

The developed prediction function can be used to determine the dry matter content with a root mean square error of 0.37% w/w in a concentration range of 5.5 to 12.4% w/w.

The prediction function developed for the determination of carbohydrate content can be used with an average error of 0.867% w/w in a concentration range of 5.4 to 36.3% w/w.

Prediction of total carbohydrate content can be performed with an average error of 0.936% w/w in a concentration range of 5.4%–36.3% w/w, while for the protein content, this average prediction error is 0.974% w/w in the concentration range of 12.9%–34.6% w/w.

Using iPLS, the root mean square error of the original prediction was improved from 81.8 kJ/100 g to 44.4 kJ/100 g. These data also demonstrate experiment to use a variable selection method for such a complex property.

All six investigated species were successfully discriminated by the LDA-supervised pattern recognition method solely based on the NIR spectra. Established estimation relationships and successful pattern recognition can be of great help to mushroom growers to choose the right species and variety.

The primary aspect of processing plants and of the processing trade is quality raw materials, because these are the base for quality products.

The FT-NIR method allows obtaining a rapid answer to this issue, and furthermore due to its versatility, also offers the possibility of in-line, on-line or at-line application.

ACKNOWLEDGMENTS

The Project was supported by the European Union and co-financed by the European Social Fund (grant agreement no. EFOP-3.6.3-VEKOP-16-2017-00005). We are grateful for the support of the Szent István University’s Doctoral School of Food Sciences.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Marietta Fodor  [1] https://orcid.org/0000-0001-8647-3201

REFERENCES

Barroetañe, C., & Toledo, C. V. (2017). The nutritional benefits of mushrooms. In I. C. F. R. Ferreira, P. Morales, & L. Barros (Eds.), Wild Plants, mushrooms and nuts: Functional food properties and applications (pp. 1–477). Chichester: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118944653

Bi, Y., Yuan, K., Xiao, W., Wu, J., Shi, C., Xia, J., ... Zhou, G. (2016). A local pre-processing method for near-infrared spectra, combined with spectral segmentation and standard normal variate transformation. Analytica Chimica Acta, 909, 30–40. https://doi.org/10.1016/j.aca.2016.01.010

Bhattacharyya, D. K., Paul, R. K., Miah, M. N., & Ahmed, K. U. (2015). Comparative study on nutritional composition of oyster mushroom (Pleurotus ostreatus Fr.). Cultivated on different sawdust substrates. Bioresearch Communications, 1(2), 93–98. Retrieved from https://pdfs.semanticscholar.org/969a/ee8182b303f50caaa503b189a2e83494b242.pdf.

Castro-Alves, V. C., & do Nascimento, J. R. O. (2016). Speeding up the extraction of mushroom polysaccharides. Food Analytical Methods, 9(9), 2429–2433. https://doi.org/10.1007/s12161-016-0406-9

Cen, H., & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. Trends in Food Science and Technology, 18(2), 72–83. https://doi.org/10.1016/j.tifs.2006.09.003

Chauhan, G., Prasad, S., Rathore, H., & Sharma, S. (2017). Nutritional profiling and value addition of products from Hypsizygus tessellatus. Food Biology, 6, 1–6.https://doi.org/10.19071/tfibiol.2017.v6.i319

Chen, H., Song, Q., Tang, G., Feng, Q., & Lin, L. (2013). The combined optimization of Savitzky-Golay smoothing and multiplicative scatter correction for FT-NIR PLS models. ISRN Spectroscopy, 2013, 1–9. https://doi.org/10.1155/2013/642190

Chen, Y., Xie, M., Zhang, H., Wang, Y., Nie, S., & Li, C. (2012). Quantification of total polysaccharides and triterpenoids in Ganoderma lucidum and Ganoderma atrum by near infrared spectroscopy and chemometrics. Food Chemistry, 135(1), 268–275. https://doi.org/10.1016/j.foodchem.2012.04.089

European Commission. (2011). Regulation (EU) No 1169/2011 of the European parliament. Official Journal of the European Union (1169), 18–63.

Ferreira, I. C. F. R., Morales, P., & Barros, L. (2017). Wild plants, mushrooms and nuts: functional food properties and applications. In I. C. F. R. Ferreira, P. Morales, & L. Barros (Eds.), Wild plants, mushrooms and nuts: Functional food properties and applications. Chichester: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118944653

Gafta. (2018). Method 10.1: Determination of sugar: Luff-Schoorl Method. January, 4.

Gezer, K., Kaygusuz, O., Herken, E. N., Dodurga, Y., Kozihaiganova, M., & Seçme, M. (2016). Evaluation of the nutritional composition of wild...
