Food Analytical Methods (2022) 15:56–66
https://doi.org/10.1007/s12161-021-02097-y

Classification of Tokaj Wines by Ultraviolet–Visible Spectroscopy

Jana Sádecká[1] · Michaela Jakubíková[1]

Received: 22 March 2021 / Accepted: 23 July 2021 / Published online: 16 August 2021

Abstract
The suitability of UV–Vis spectrometry was evaluated for the classification of undiluted and diluted Slovak Tokaj wine samples according to style (essences, Tokaj selection, varietal and other wines), grade (quantity of cibebas), and variety (Furmint, Lipovina, and Muškát žltý) using principal component analysis (PCA), variable selection (VS), linear discriminant analysis (LDA), general discriminant analysis (GDA), and support vector machine (SVM). The individual groups of Tokaj wines differed in their production process, the quantity of cibebas used in their production, and the accorded Protected Designation of Origin status, all of which determined their price. In general, it was found that better classification was obtained based on the UV–Vis spectra of the diluted samples, VS was a more suitable algorithm for reducing the number of variables than PCA, and finally, LDA/GDA was preferred over SVM. The best total correct classification (100%) was obtained using diluted wines and VS-GDA method. To achieve this result, 45, 31, and 10 variables were needed for classification by style, grade, and variety, respectively. Thus, UV–Vis spectrometry combined with chemometrics can be widely exploited for quality control and authentication of Tokaj wines because of a relative simplicity, short-time analysis, and without considerable financial expenses.

Keywords Beverage · Wine · Ultraviolet–visible spectroscopy · Chemometrics

Introduction

UV–Vis spectrometry is a fast, simple, and cheap technique which has found wide acceptance in food chemistry. In the wine industry and research, well-known applications include the specification of chromatic characteristics and the determination of total phenolic content by the so-called Folin-Ciocalteu assay (OIV 2016). Besides, there are several review articles dealing with other applications of the UV–Vis spectroscopic technique, including the determination of phenolic compounds in grape and wine (Cozzolino 2015), the simultaneous prediction of several wine chemical properties (Yu et al. 2017), the determination of geographical origin of wines (Uríčková and Sádecká 2015), and the discrimination and authentication of wines (Chandra et al. 2017; Yu et al. 2017). Some recent applications include the determination of total acid, total sugar, and alcohol in wines based on visible and near infrared spectra (Hu et al. 2018) and the determination of methyl cellulose precipitable tannins, anthocyanins, total phenols, and color density during the fermentation as well as in finished red wines, using only UV–Vis spectra (Aleixandre-Tudo et al. 2018); the geographical classification of Cabernet Sauvignon red wines by data fusion of UV–Vis and synchronous fluorescence spectra (Tan et al. 2016); varietal and vintage classifications of red and white wines by combining UV–Vis spectra with either color properties, pH, and total anthocyanin contents (Sen and Tokatli 2016) or the total phenolic index and color characteristics (Philippidis et al. 2017); the discrimination of brands of red wines using only UV–Vis spectra (Liu et al. 2018); and the prediction of the ageing time of Portuguese fortified wines, based on the volatile, polyphenols, organic acid composition, and the UV–Vis spectral data, which were processed individually (Rendall et al. 2017) as well as simultaneously (Campos et al. 2017) and the differentiation between the seven levels of ageing of Spanish wines using two different fingerprint ranges in infrared and one range in the visible region (Ferreiro-González et al. 2019). Besides, the effect of path length on the standard error of UV, VIS, and NIR calibration models to predict phenolic compounds

✉ Michaela Jakubíková
michaela.jakubikova@stuba.sk

1 Faculty of Chemical and Food Technology, Institute of Analytical Chemistry, Slovak University of Technology, Radlinského 9, 812 37 Bratislava, Slovak Republic

in both red and white wines was evaluated; no statistically significant differences were observed. However, the use of different path lengths can lead to opposite conclusions in classification or discrimination issues (Molla et al. 2017).

Tokaj region is located on both sides of the border between Slovakia and Hungary. Slovak Tokaj region is demarcated by legislation—only wine made in the designated region from the grapes growing in that region can legally be designated Tokaj. Slovak Tokaj wines are made exclusively from white wine grape varieties Furmint, Lipovina, and Muškat žltý, which are used to produce single-varietal a semi-sweet or a dry still wines as well as botrytized wines. Single-variety wine is not exclusively made from a single variety but may contain 15% of the remaining two varieties, e.g., Furmint is produced by alcoholic fermentation of the grape variety Furmint with the addition of grapes of the varieties Lipovina and Muškát žltý, together a maximum of 15%. The production of botrytized wines is closely linked to a special technology that requires adding the exact amount of botrytized grapes (cibebas) to a set volume of wine and then maturing of the wine in oak barrels in cellars for a period of several years (Figueiredo-Gonzalez et al. 2013; Law No. 313/2009 on viticulture and wine; Magyar 2011). The amount of added cibebas is one of the factors determining the price of wines. The most expensive is the essence because it is made from cibebas only. Another type of botrytized wine is a Tokaj selection, which can be 6-, 5-, and 4-putňový, depending on the amount of cibebas added. All these wines were accorded Protected Designation of Origin status by the European Commission (Commission Regulation (EU) No 401/2010). Unprotected 3- and 2-putňový wines are also produced from time to time. Wine pomace, which remains after the production of Tokaj selection and contains many extractive compounds and sugar, is used for the production of Tokajský fordítáš—Tokaj wine or good quality must is poured onto wine pomace and allowed to ferment (Law No. 313/2009 on viticulture and wine). In unfavorable years, when not enough cibebas are formed on bunches of grapes, the raw material is intended for the production of `Samorodné` sweet wine (Magyar 2011). The other two types of wine, `Samorodné` dry wine and semi-dry wine Víno kráľovnej Alžbety, are made from grapes without cibebas.

Information on the composition of Slovak Tokaj wines is rare. The distribution of the Ca, Mg, Rb, Sr, Ba, and V elements in white wines allowed the classification of wines from different Slovak vineyard regions, with the Tokaj region being the most separated (Koreňovská and Suhaj 2005). Ballová et al. (2016) studied total phenolic content and antioxidant activity and found that these two properties were decreased in the following order: 6-, 5-, 4-, 3-putňový, Muškát žltý, Furmint, and Lipovina. On the other hand, high-resolution 1H NMR spectroscopy showed similar relative integral intensities for 6-, 5-, and 4-putňový wines as well as essence, corresponding to the phenolic compounds (Mazur et al. 2015). Only recently, essences and 6-, 5-, and 4-putňový wines (vintages 2000–2015) (Sádecká et al. 2018) as well as three Tokaj varietal wines (Sádecká et al. 2020) were distinguished using linear discriminant analysis (LDA), based on emission and synchronous fluorescence spectra. A highly sophisticated technique, comprehensive two-dimensional gas chromatography coupled to high-resolution time-of-flight mass spectrometry, has made it possible to distinguish Tokaj varietal, Samorodné dry, and Tokaj selection wines using the profile of volatile organic compounds (Furdíková et al. 2020). More data can be found about Hungarian botrytized wines, but unfortunately, because of a small number of samples, the studies never used chemometrics (Magyar 2011; Pour Nikfardjam et al. 2003, 2006).

The aim of this study was to evaluate the use of UV–Vis spectroscopy combined with chemometric methods to discriminate between different Slovak Tokaj wines according to the style, grade (the amount of cibebas added), and variety.

Materials and Methods

UV–Vis Spectra

UV 1800 spectrophotometer (Shimadzu, Japan) equipped with tungsten iodine and deuterium lamp was used to record absorption spectra. The spectrophotometer operated in the wavelength range of 200 – 800 nm with accuracy of $\pm$ 0.1 nm for the D2 peak (656.1 nm) and $\pm$ 0.3 nm for the full range. The scanning speed was 200 nm•min^{-1}. Software UV PROBE 2.33 was used for acquisition and processing of spectra. The three UV–Vis spectra for each sample were averaged, and the mean spectra were used to chemometric analysis.

Samples

A total of 77 Slovak Tokaj wine samples were collected from producers in the Tokaj region and from local stores. The wines were divided by style into four groups: Tokaj selection wines (abbreviated as P, $n=49$), essences (E, $n=4$), varietal wines (V, $n=14$), and other wines (O, $n=10$) (Table 1). Most of the samples, 49, were Tokaj selection wines of varying grades from 2 to 6, depending on the quantity of cibebas added. Two of Tokaj selection wines were 2-putňový (P2), eight were 3-putňový (P3), ten were 4-putňový (P4), fourteen were 5-putňový (P5), and finally fifteen were 6-putňový (P6) botrytized wines. The oldest botrytized wine was from 1972 and the youngest from 2016. Four samples were Tokaj essences (E) (vintages from 1999 to 2009). A group of varietal wines (V) consisted of three samples of

Table 1 Tokaj wines used for the classification study

Wine style		Number of samples	Vintage
Essences (E)		4	1999–2009
Tokaj selection (P)	Grade:		
	2-putňový (P2)	2	1989–1990
	3-putňový (P3)	8	1988–2009
	4-putňový (P4)	10	1993–2016
	5-putňový (P5)	14	1972–2004
	6-putňový (P6)	15	1972–2006
Varietal (V)	Lipovina (L)	3	2015
	Furmint (F)	6	2012–2014
	Muškat žltý (M)	5	2012–2016
Other (O)	Víno kráľovnej Alžbety (KrAlz)	1	2016
	Tokajský fordítáš (TF)	1	2011
	Samorodné sweet wine (SSd)	2	2006
	Samorodné dry wine (SSch)	6	2009–2016

Lipovina (L), six samples of Furmint (F), and five samples of Muškat žltý (M) (vintages from 2012 to 2016). The group of other wines (O) (ten samples) was very diverse, consisted of Víno kráľovnej Alžbety (KrAlž), Tokajský fordítáš (TF), two samples of Samorodné sweet wine (SSd), and six samples of Samorodné dry wine (SSch) (vintages from 2006 to 2016). The wines were stored in the dark, at room temperature until its analysis. The UV–Vis spectrum of each sample was registered immediately after opening each bottle. Bulk and diluted (1:100, v/v, ultrapure water, electrical resistivity 18.2 M Ω cm) wine samples were analyzed.

Chemometrics

Multivariate data analysis, including unsupervised method, principal component analysis (PCA), and three supervised methods, linear discriminant analysis (LDA), general discriminant analysis (GDA), and support vector machine (SVM), was carried out by STATISTICA version 12 (StatSoft, USA, 2017).

First, spectral datasets were pre-processed by a mean-centering algorithm. PCA was applied to spectral datasets to reduce the number of original correlated variables. PCA resulted in new uncorrelated variables, called principal components (PCs), whose scores were used as input data to the LDA in the next step (Martelo-Vidal et al. 2013; Sádecká et al. 2018; Tan et al. 2016). The number of PCs was tentatively estimated based on the total explained variance (> 90%) and loading profiles and was refined during PCA-LDA model validation by ten-fold cross-validation method (Arlot and Celisse 2010; Sádecká et al. 2018). It was the number of PCs for which another added component did not significantly increase the overall correct classification.

Second, the pre-processed spectral datasets were used to build different types of support vector machine (SVM) models, including linear, polynomial, sigmoid, and radial basis function (RBF) kernels. While LDA assumes that data is normally distributed and all groups are identically distributed, SVM makes no assumptions about the data at all. SVM is also suitable for small sample sets and some non-linear problems; thus, it is a very flexible method (Martelo-Vidal et al. 2013). Depending on the type of kernel, specific parameters were optimized by combining ten-fold cross-validation with a grid search. The specific parameters were degree, gamma, and coefficient for polynomial kernel, gamma for RBF kernel, and gamma and coefficient for sigmoid kernel. PCA-based approaches associated with SVM have been proposed as highly effective for classifying samples (Dankowska and Kowalewski 2019; Yang et al. 2017). Therefore, SVM models were also created using the same PCs as in the LDA case and ten-fold cross-validation method.

The removal of useless variables can enhance the accuracy and decreases the computation time (Ikram and Cherukuri 2017). Therefore, variable selection based on chi-square was used to find the features that best separate the groups within each dataset, and finally, GDA and SVM models were built by using the selected variables.

The typical development of a chemometric procedure started with data pre-processing, and then all data was used to create a well-defined model, selecting variables, setting parameters, and performing other development steps. Then the model was validated. In the case of classification by style and grade, a stratified four-fold cross-validation was used (Vabalas et al. 2019). The calibration and validation sets were created so that approximately one-fourth of the samples were included in the validation sets and the remaining three-fourth of the samples in the calibration sets. The samples were selected in approximately the same proportions as they appeared in the original set. Validation process was repeated four times. In each fold, a different one-fourth of the samples were used for validation. In this way, at the end, all the samples were used for both calibration and validation. The final performance of the model was then calculated as a mean of classification performances in each of the four validation folds. Example of a four-fold cross-validation data split for classification of wines according to style is shown in Table 1S (Supplementary Information). In the case of the classification of wines by variety, the leave-one-out cross-validation (LOOCV) method was used because the number of varietal wines was small. In the LOOCV for 14 samples, each calibration set was created by taking all the samples except one, and the validation set was the sample left out. This procedure was repeated 14 times, and then the final performance of the model was determined as the average of 14 steps.

Results and Discussion

UV–Vis Spectra

Figure 1 shows the raw average UV–Vis spectra for Tokaj selection (P2-P6), varietal (F, L, M), and other (KrAlz, TF, SSd, SSch) wines. The absorbance of the undiluted samples was greater than 4 in the wavelength range of 200–350 nm and then gradually decreased without significant features to a negligible value at a wavelength longer than 600 nm (Fig. 1a, b, c). Two maxima around 265–280 nm and 325 nm were observed for the diluted samples, and the absorbance was small at wavelengths near 400 nm and longer (Fig. 1d, e, f). Spectra were similar to those previously described and assigned to aromatic compounds as hydroxybenzoic acids (250–300 nm), hydroxycinnamic acids (230–245 nm and 310–330 nm), catechins (280 nm), and flavonols (250–270 and 350–390 nm) (Aleixandre-Tudo et al. 2018; Martelo-Vidal et al. 2013; Molla et al. 2017; Tan et al. 2016; Yu et al. 2017). Although the spectral profiles of the samples from the different groups were similar, they often differed in absorbance at the same wavelength. Subsequently, the average absorbance of the Tokaj selection wines was arranged in the order of P6 > P5 > P4 = P3 = P2 in the range of 370–600 nm and 200–235 nm for the undiluted and diluted samples, respectively (Fig. 1 a and d). The exact positions of the maximum for diluted P6, P5, P4, P3, and P2 samples were observed at 280, 278, 274, 272, and 280 nm, respectively, with the absorbance decreasing from P6 to P2. A similar decrease was also seen for shoulder at 325 nm. The UV–Vis spectra of diluted varietal wines (Fig. 1e) were more similar to each other compared to those of Tokaj selection wines not only in the position

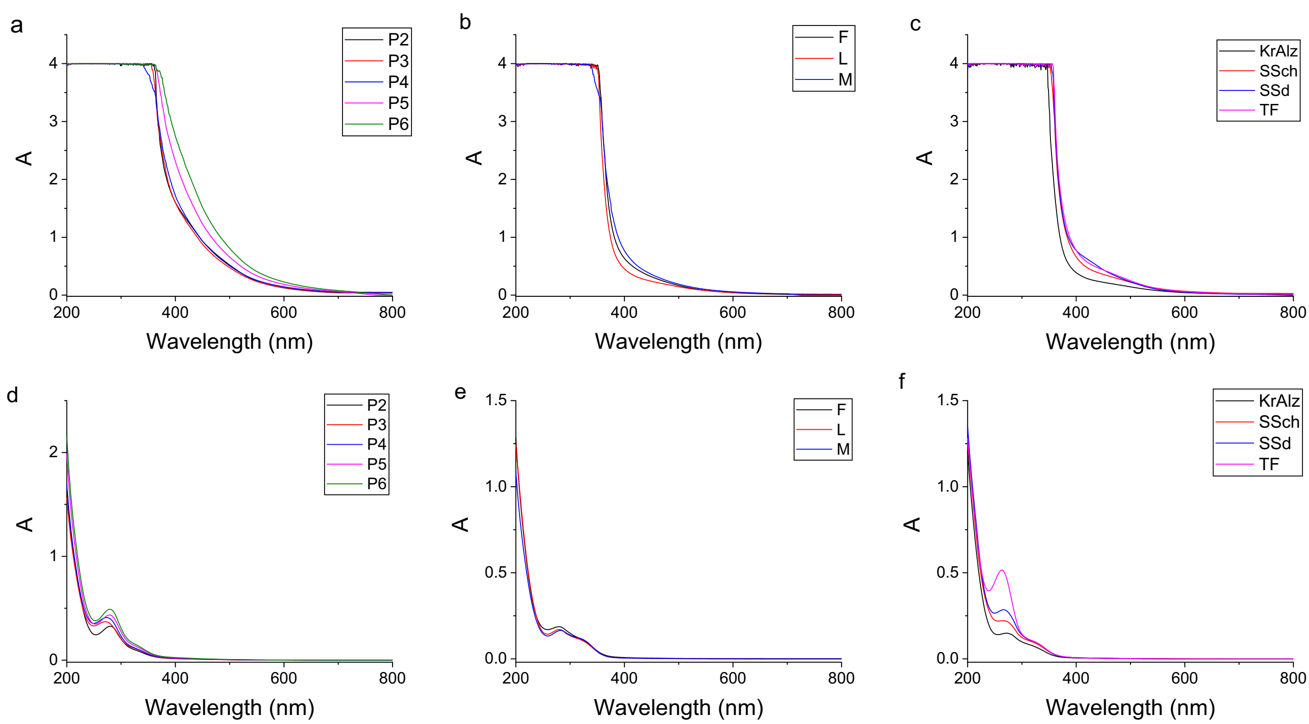

Fig. 1 The average UV–Vis absorption spectra of undiluted (**a**, **b**, **c**) and diluted (**d**, **e**, **f**) Tokaj wines (Tokaj selection wines, 2-putňový, P2; 3-putňový, P3; 4-putňový, P4; 5-putňový, P5; and 6-putňový, P6; varietal wines, Furmint, F; Lipovina, L; and Muškát žltý, M; and other wines, Víno kráľovnej Alžbety, Kr.Alz.; Tokajský fordítáš, TF; Samorodné sweet wine, SSd; and Samorodné dry wine, SSch)

of maximum (280–282 nm) and shoulder (325 nm), but also in the absorbance values especially in the range of 300–400 nm.

Diluted KrAlz, SSch, and SSd from the group of other wines showed similar absorbance as varietal wines, but absorption maxima appeared at lower wavelengths—272 nm (KrAlz), 269 nm (SSch), and 266 nm (SSd) (Fig. 1f). Diluted TF was best distinguished from all other types of wines due to the typical band with a maximum at 263 nm (Fig. 1f).

An overview of the UV–Vis spectra of diluted samples suggests a possible relationship between the amount of cibebas used in winemaking and the shape of the spectra. The largest amount of cibebas is used in the production of essence and P6, which corresponds to the greatest absorbance and similarity in the spectra. Tokaj forditáš is made from grape/cibeba pomace remaining after the production of the Tokaj selection wines. Its absorbance was similar to that of essences and P6 wine, but the absorption band was shifted towards lower wavelengths. A decrease in the amount of cibebas leads to a decrease in absorbance from P6 to P2 and then to SSd, which is produced in climatically unfavorable years for the formation of cibebas. Samorodné dry, semi-dry, and varietal wines are made from grapes without cibebas and are therefore characterized by low absorbance. These observations are consistent with reports of the effect of *Botrytis cinerea* on the global phenolic compound content in grapes and the associated differences between botrytized and non-botrytized grape (cibeba)/wine, e.g., increasing the amount of catechin, epicatechin, epicatechin gallate (Carbajal-Ida et al. 2016), flavan-3-ols, furfuraldehyde (Figueiredo-Gonzalez et al. 2013; Furdíková et al. 2020),p-coumaric acid dimer, vanillic acid, syringic acid (Zimdars et al. 2017), total phenols (Ballová et al. 2016; Magyar 2011; Pour Nikfardjam et al. 2003; Pour Nikfardjam et al. 2006), flavonoid glycosides, and flavanones (Magyar 2011) with the level of botrytization. These components, together with others, can contribute to the observed absorption of UV–Vis light by Tokaj wines.

Classification According to Style

For the purposes of this part, wines have been divided into four styles: essences (E), Tokaj selection wines (P), varietal wines (V), and other wines (O).

PCA for undiluted samples resulted in four principal components describing 99.9% of total variability, with the PC1 accounting for 89.1% (PC2, 6.2%; PC3, 3.4%; PC4, 1.2%). The loadings are shown in Fig. 1S (Supplementary Information). The PC1 vs. PC2 score plot (Fig. 2a) shows that the samples were partially separated by style based on PC1, with the Tokaj selection wines on the right and other wines on the left. Unfortunately, the essence group completely coincided with the Tokaj selection wine group not only in PC1 but also in PC2. Varietal and other wines have been partially distributed on the basis of PC2. Seven PCs describing 99.6% of total variability (PC1, 82.6%; PC2,

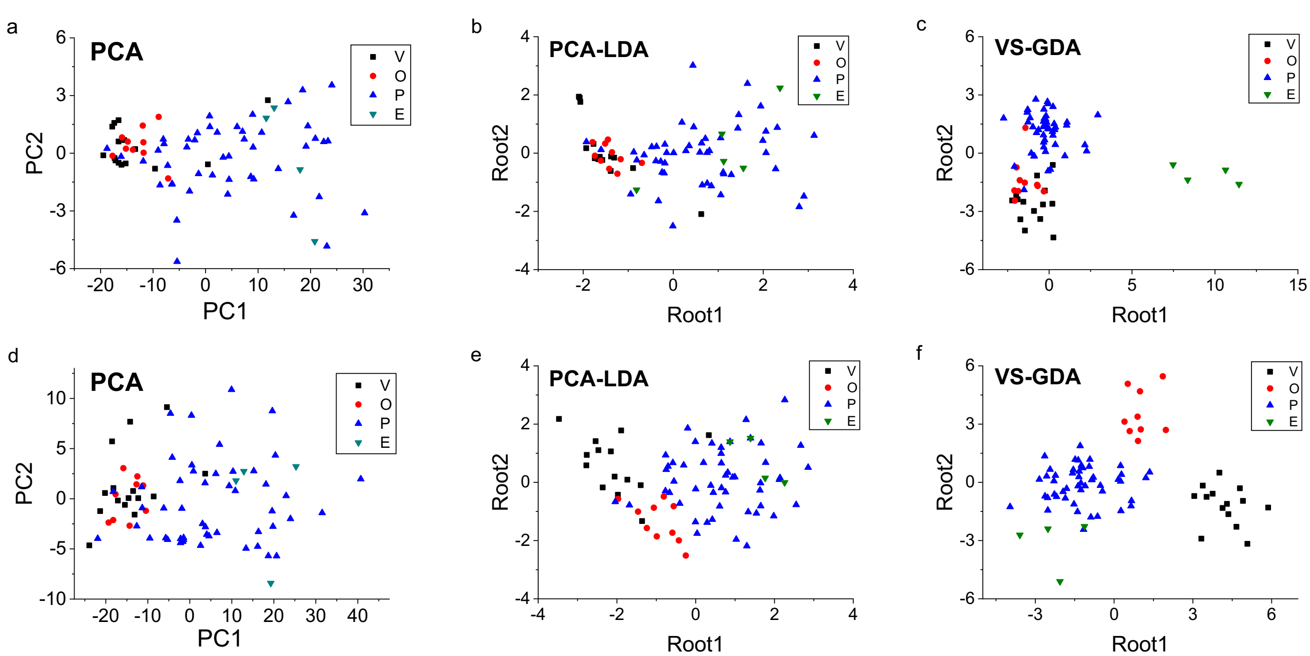

Fig. 2 Chemometric results for classification according to style on undiluted (**a**, **b**, **c**) and diluted (**d**, **e**, **f**) Tokaj wines (essences, E; Tokaj selection wines, P; varietal wines, V; and other wines, O). PCA, principal component analysis; PCA-LDA, principal component analysis-linear discriminant analysis; VS-GDA, variable selection-general discriminant analysis

6.1%; PC3, 4.3%; PC4, 2.7%; PC5, 1.9%; PC6, 1.2%; PC7, 0.8%) were found for diluted wines. The loadings are shown in Fig. 1S (Supplementary Information). The PC1 vs. PC2 score plot (Fig. 2d) looked very similar to that of undiluted samples (Fig. 2a). Again, the Tokaj selection wines were partially separated from the others on a PC1 basis, with the essences overlapping with the Tokaj selection wines in both the PC1 and PC2.

Although the separation of the diluted samples was better than the undiluted samples, complete separation was not achieved. Therefore, in the next step, LDA was used, based on the first PCs, which resulted in three discriminatory functions (Root). Scatter plot of canonical scores for Root1 vs. Root2 (Fig. 2b) shows a similar distribution of samples as in the PC score graph for undiluted samples. On the contrary, a better separation of Tokaj selection wines and varietal and other wines was observed for the diluted samples (Fig. 2e). Root1 mostly discriminated between Tokaj selection wines and the group of varietal and other wines, while Root2 discriminated between varietal and other wines. Overall, the LDA allowed better discrimination compared to PCA: 79.2% based on 4 PCs for undiluted samples and 84.4% based on 7 PCs for diluted samples (Table 2). SVM applied to PC scores generally yielded worse results compared to PCA-LDA, with linear function providing the best results in SVM. The overall correct classification based on PCA-SVM was very similar to that achieved by applying SVM to the whole spectral range (Table 2).

In the next step, a feature selection algorithm was used to sort the variables according to the decreasing chi-square value for which values ranged from 90 to 63 (undiluted samples) and from 85 to 30 (diluted samples) were obtained. From these values, it was not possible to clearly identify the appropriate number of variables, so different numbers of variables were gradually used in both GDA and SVM, resulting in an overall correct classification as shown in Fig. 2S (Supplementary Information). For undiluted samples, 45 variables with the largest chi-square value (> 70) resulted in 96.1% correct classification using GDA (Table 2). Two Tokaj selection wines were assigned to the wrong group, the varietal or other wine group. One varietal wine was misclassified as Tokaj selection. All essences and all other wines were correctly classified. Comparison of the score plots showed that, unlike the PCA-LDA, essences were well separated by Root1 and that Tokaj selection wines were separated by Root2 in VS-GDA (Fig. 2c). Using the same 45 variables in SVM resulted in a worse classification compared to GDA with the best result (81.8%) using linear function in SVM (Table 2). Figure 2S (Supplementary Information) shows that the addition of other variables did not improve the classification in either GDA or SVM. For diluted samples, again 45 variables (chi-square > 67) were selected that resulted in 100 and 79.2% classification using GDA and linear SVM, respectively (Table 2). Other functions in SVM gave worse results. Scatter plot for Root1 vs. Root2 visualized a good differentiation between four wine styles, where varietal wines were best differentiated according to Root1 and other wines were best resolved according to Root2 (Fig. 2f). Diluted essences were much more similar to Tokaj selection wines than undiluted essences.

Comparing the results in the Table 2 shows that selecting variables before GDA was a better strategy than PCA combined with LDA. This was mainly seen for the diluted samples. In SVM, the result was less dependent on the strategy used.

Most relevant wavelengths after variable selection (undiluted samples: 400 – 413, 418, 451, 453, 470, 475, 479, 484, 503, 509, 511, 517, 520, 525, 552, 562 – 566, 568 – 573, 584, 586, 588, 589, 591, and 592 nm; diluted samples: 287, 292, 294, 295, and 365 – 405 nm) and average spectra are shown in Fig. 3S (Supplementary Information). The Vis spectral regions can be related to browning pigments in grape and wines resulted from enzymatic and non-enzymatic reactions taking place during grape dehydration and ageing of wines (Figueiredo-Gonzalez et al. 2013; Ferreiro-González et al. 2019). Enzymatic browning occurs in grape must; it is largely correlated with the content of caftaric and coutaric acids, promoted by flavanols.

Table 2 Chemometric results for classification according to style

	Number of variables or PC	LDA (GDA)	SVM Linear	SVM Polynomial	SVM RBF	SVM Sigmoid
Undiluted samples						
Whole range 400 – 600 nm	201		80.5	63.6	76.6	76.6
PCA	4	79.2	79.2	63.6	67.5	63.6
Variable selection	45	(96.1)	81.8	63.6	76.6	75.3
Diluted samples						
Whole range 230 – 479 nm	250		85.7	63.6	75.3	75.3
PCA	7	84.4	80.6	63.6	75.3	68.8
Variable selection	45	(100)	79.2	63.6	76.6	76.6

Non-enzymatic browning can arise in both grape must and wine through several pathways, including oxidation and polymerization related to caffeic, caftaric, and gallic acids, catechin, epicatechin, and carbohydrates (Figueiredo-Gonzalez et al. 2013). Volatile organic compounds (VOCs), phenylmethanol, 2-phenylethanol, 2-metoxyphenol, furfural, 5-methylfurfural, and guaiacol, have been reported by Furdikova et al. (2020) to be responsible for the greatest differences between Tokaj selection, Tokajské samorodné dry, and Tokaj varietal wines. The VOCs together with phenolic compounds contribute to absorption in the UV region. Pour Nikfardjam et al. (2003) compared Hungarian Tokaj wines, Eszencia, Tokaji Aszú, Samorodni, Forditás, and varietal wines, observing differences in the content of total phenols, gallic, caftaric, coutaric, and p-coumaric acids, catechin, and epicatechin. The highest contents were as follows: coutaric acid, p-coumaric acid, and total phenols in Eszencia, gallic acid in Samorodni wine, catechin and epicatechin in Forditás, and caftaric acid in varietal wines.

Classification of Tokaj Selection According to Grade

In this part, Tokaj selection wines were classified according to the amount of cibebas used in their production. Applying PCA to UV spectra of Tokaj selection wines, four PCs (PC1, 96.9%; PC2, 2.4%; PC3, 0.5%; PC4, 0.1%; total, 99.9%) and seven PCs (PC1, 80.4%; PC2, 7.8%; PC3, 5.3%; PC4, 3.2%; PC5, 2.1%; PC6, 0.7%; PC7, 0.2%; total, 99.7%) were found for undiluted and diluted wines, respectively. The loadings are given in Fig. 1S (Supplementary Information). The PCs scores were subsequently used in LDA and SVM. The PC1 vs. PC2 score plots (Fig. 3 a and d) show that the separation was not clear, some overlap was observed between the groups, e.g., P6 with P5, and a large dispersion of score was observed mainly for P5 and P6 group. As can be seen from the Table 3 and Fig. 3 b and e, PCA-LDA failed to discriminate wines by grade as the classification rate was below 50% for both undiluted and diluted samples. A smaller dispersion of score in the groups was observed for undiluted samples, e.g., the group of undiluted P2 samples is relatively homogeneous compared to the diluted ones; however, P5 and P6 samples were still too dispersed to avoid overlap. In addition, the classification rate below 50% was also obtained by applying SVM to either PCs or UV spectra over the whole spectral ranges (Table 3).

In the next step, the most important variables were selected in a similar way as it was done for different wine styles; the variables were arranged according to the decreasing chi-square value, and then GDA/SVM models were calculated with different numbers of variables, resulting in a total correct classification shown in Fig. 2S. Chi-square value range was from 79 to 51 for undiluted samples, and it was from 71 to 25 for diluted samples. Thirty-one variables with chi-square value > 70 were selected for undiluted samples, and 31 variables with chi-square value > 56 were selected for diluted samples. Most significant wavelengths

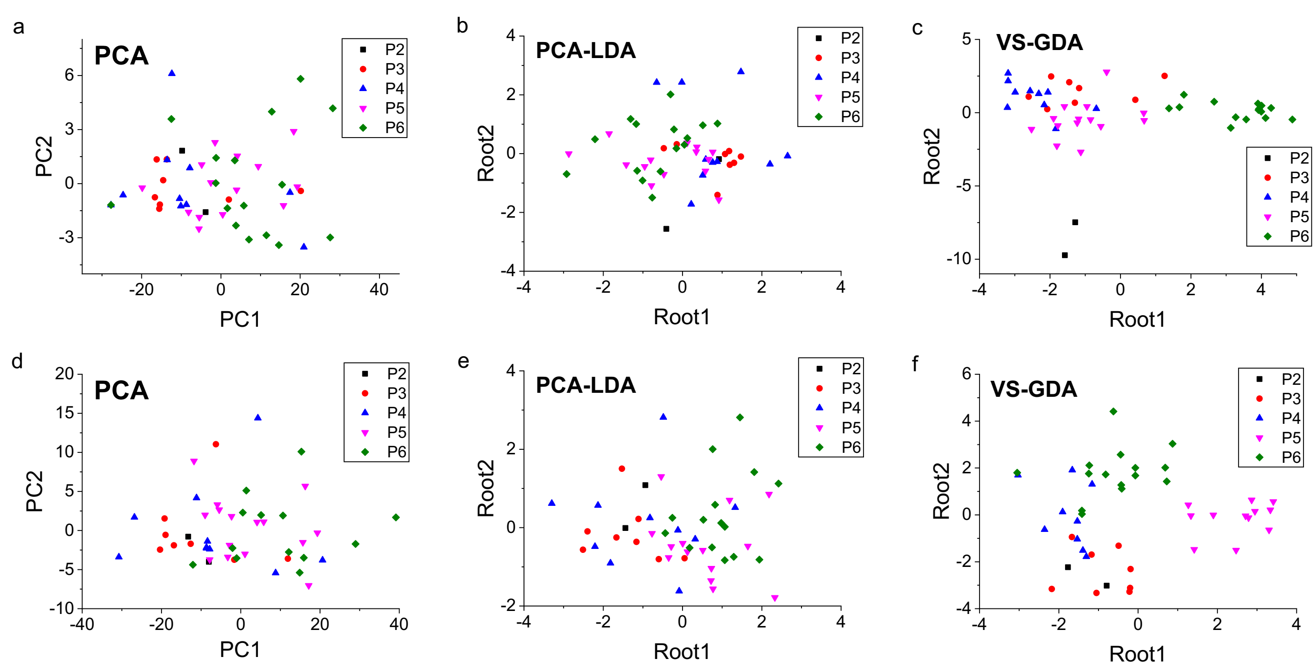

Fig. 3 Chemometric results for classification of Tokaj selection according to grade on undiluted (**a**, **b**, **c**) and diluted (**d**, **e**, **f**) Tokaj wines (2-putňový, P2; 3-putňový, P3; 4-putňový, P4; 5-putňový, P5; and 6-putňový, P6 wines). PCA, principal component analysis; PCA-LDA, principal component analysis-linear discriminant analysis; VS-GDA, variable selection-general discriminant analysis

Table 3 Chemometric results for classification of Tokaj selection according to grade

	Number of variables or PC	LDA (GDA)	SVM Linear	SVM Polynomial	SVM RBF	SVM Sigmoid
Undiluted samples						
Whole range 400 – 600 nm	201		48.9	28.5	42.8	40.8
PCA	4	46.9	46.9	30.6	44.9	32.6
Variable selection	31	(97.9)	44.9	30.6	42.9	44.9
Diluted samples						
Whole range 230 – 479 nm	250		49.0	28.6	28.6	28.6
PCA	7	48.9	28.6	28.6	28.6	28.6
Variable selection	31	(100)	51.0	28.6	28.6	28.6

were 415, 416, 421 – 423, 425 – 430, 433 – 440, 452, 465, 466, 467, 468, 469, 508, 519, 522 – 524, and 526 nm for undiluted samples, and these were 243, 245, 246, 352 – 363, 367, 368, 379, 381 – 385, 450 – 454, 456, and 457 nm for diluted samples (Fig. 3S). SVM models calculated using 31 variables with the highest value of the chi-square achieved nearly equally poor quality of classification (< 50%) as those for full spectral ranges (Table 3, Fig. 2S).

Although the variable selection algorithm did not improve the classification using SVM, it was very useful in combination with GDA. For undiluted samples, 31 variables allowed 97.9% of samples to be classified correctly (Table 3), with P2, P3, P5, and P6 samples all classified correctly and only one sample of P5 classified as P3. A better classification compared to PCA and PCA-LDA was also clearly visible in the score plot for VS-GDA where all P6 samples had positive score for Root1, P5 samples had small and mostly negative score for Root1 and Root2, large negative Root1 and Root2 were observed for P2 samples, while P3 and P4 samples were most similar in Root1 and Root2 (Fig. 3c).

The best classification (100%) of Tokaj selection wines according to grade was achieved for diluted samples by using 31 variables in GDA. Figure 3 f shows Root1 vs. Root2 score plot in which the P6 samples were observed on the right and the P5 samples were located on the top, however, partially overlapped with P4 samples. In addition, similarity was observed between P2 and P3 samples. Although a clear discrimination between the five groups was not observed, the grouping of wine samples according to grade was better compared to PCA and PCA-LDA scatter plots.

The significant region for the classification of undiluted samples is the visible one, particularly the wavelengths at about 420, 450, and 520 nm, which are related to browning index, xanthylium pigments, and polymeric pigment color, respectively. *B. cinerea* produces the enzyme laccase, which oxidizes a large group of phenolic compounds to quinones, followed by polymerization of the quinones to form brown components. Therefore, the optimally botrytized grape berries (cibebas) are brown (Figueiredo-Gonzalez et al. 2013). Tokaj selection wine shall mature at least 2 years in wooden cask, and the ageing process also affects the absorption in the VIS range of 380–450 nm (Ferreiro-González et al. 2019). The characteristic absorption band of xanthylium derivatives occurs in the range 440–460 nm, part of which is also relevant for the classification of diluted samples. Xanthylium derivatives are formed by the reaction of flavanols with furfural and 5-hydroxymethyl, which are degradation products of sugars (Bührle et al. 2017). Note that Tokaj selection wines differ in residual natural sugar content, which decreases from P6 to P2, from 150 to 60 g/L. Finally, significant features at lower wavelengths (352–385 nm) can be related to the flavonol group, which exhibits an absorption band at 360 nm. Flavonols as myricetin, quercetin, kaempherol, kaempherol-*O*-gallate, and quercetin-*O*-gallate have been identified in Hungarian Tokaj aszu wines (Kovács et al. 2004), and myricetin, quercetin-3 glucoside, and kaempferol-3 glucoside have been found in botrytized Chenin blanc grapes (Carbajal-Ida et al. 2016).

Classification According to Variety

In this part, the models for the classification of three varieties of wines (Furmint (F), Lipovina (L), and Muškát žltý (M)) were developed. Score plots of the first two PCs from PCA calculated for the undiluted and diluted samples (Fig. 4a and d) show that the L, M, and F samples were scattered in two, three, and four quadrants, respectively, making it impossible to classify wine varieties based on PC1 and PC2 only. Therefore, the scores of the first five PCs describing 99.9% (PC1, 94.1%; PC2, 3.4%; PC3, 1.6%; PC4, 0.6%; PC5, 0.2%) and 99.7% (PC1, 90.5%; PC2, 6.6%; PC3, 1.9%; PC4, 0.4%; PC5, 0.2%) of total variability were used for the undiluted and diluted samples, respectively, as input data to LDA and SVM in the next step. The PC loadings are shown in Fig. 1S (Supplementary Information). Using PCA-LDA, undiluted F showing a negative score in Root1 were separated from the other two varieties having a positive Root1 score. Root2 discriminated mainly between M and L, as M and L showed

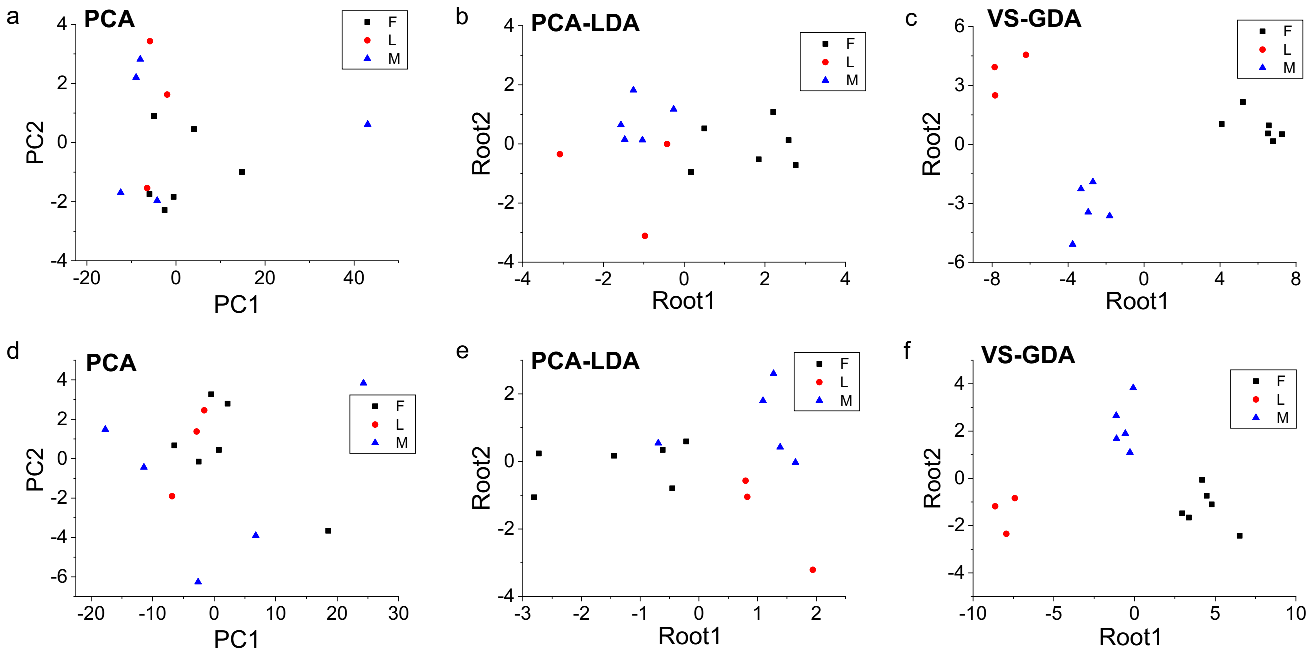

Fig. 4 Chemometric results for classification according to variety on undiluted (**a**, **b**, **c**) and diluted (**d**, **e**, **f**) Tokaj wines (Furmint, F; Lipovina, L; and Muškát žltý, M). PCA, principal component analysis; PCA-LDA, principal component analysis-linear discriminant analysis; VS-GDA, variable selection-general discriminant analysis

positive and negative scores, respectively, but for F, both positive and negative scores were observed (Fig. 4b). Total correct classification was 92.8% (Table 4), with all samples classified correctly except for one L sample, which was classified as M wine. The same percentage of total correct classification (92.8%) was achieved by applying PCA-LDA to the diluted samples, and similar pattern observed for undiluted samples was repeated with rotation around the Root2 axis for diluted ones in Fig. 4e. However, in this case, one sample of variety M was incorrectly classified as belonging to variety F. This is clearly seen in the Root 1 vs. Root2 score plot, where one M sample is located in the same quadrant as most of the F samples (Fig. 4e). While only one sample was incorrectly classified using PCA-LDA, two samples were incorrectly assigned by applying a linear SVM to the PC scores. One sample L and one sample M were both assigned to group F, resulting in 85.7% total correct classification for both undiluted and diluted wines. The same two samples together with one more M sample were classified into group F by applying either a RBF SVM to the PC scores or a linear SVM over the whole spectral ranges, resulting in 78.5% total correct classification. The same result was also

Table 4 Chemometric results for classification according to variety

	Number of variables or PC	LDA (GDA)	SVM Linear	SVM Polynomial	SVM RBF	SVM Sigmoid
Undiluted samples						
Whole range 380 – 580 nm	201		78.5	42.9	64.2	57.1
PCA	5	92.8	85.7	50.0	78.5	64.3
Variable selection	11	(100)	57.1	42.9	57.1	57.1
Variable selection	20		71.4	42.9	57.1	57.1
Variable selection	122		78.5	42.9	57.1	57.1
Diluted samples						
Whole range 230 – 360 nm	131		78.5	42.9	57.1	57.1
PCA	5	92.8	85.7	42.9	78.5	78.5
Variable selection	10	(100)	50.0	42.9	42.9	42.9
Variable selection	20		57.1	42.9	42.9	42.9
Variable selection	52		78.5	42.9	57.1	57.1

obtained for the diluted samples evaluated by PCA-sigmoid SVM (Table 4).

After calculating the relative importance of the variables, the generated variable subsets were used to build the GDA and SVM models, yielding the total correct classifications shown in Fig. 2S. Chi-square value ranges were 8 – 16 and 8 – 20 for undiluted and diluted samples, respectively. Using GDA on undiluted samples, a subset of variables was generated containing those 11 variables (398, 400, 403, 407, 412, 415, 422, 430, 443, 448, and 464 nm, Fig. 3S) that achieved the chi-square value > 13 and resulted in 100% total correct classification (Table 4). From the scatterplot for the two discriminant functions, it was clear that the first discriminant function Root1 mostly discriminated between F variety and the two others. In addition, Root2 separated well M variety from the two others (Fig. 4c). Using the same 11 variables in SVM resulted in a worse classification compared to GDA, with a total correct classification of 57.1% provided by linear, RBF, and sigmoid kernels in SVM (Table 4). Figure 2S shows that the addition of other variables did not significantly improve the classification except for the linear kernel, for which a higher classification, 78.5%, was achieved using 122 variables. However, considering all variables (whole spectral range), 78.5% classification was again obtained with the same incorrectly classified samples (one L and two M as F).

For the GDA of diluted samples, 10 wavelengths (320, 322, 325 – 327, 333, and 338 – 341 nm, Fig. 3S3S) characterized by the chi-square value > 19.6 provided total correct classification of 100% (Table 4). Three well-defined clusters were formed in Fig. 4f in which Root1 and Root2 separated F variety and M variety, respectively, similarly to undiluted samples. Another similarity was observed when linear SVM was applied, in which 10, 52, and 131 (whole range) variables allowed 50.0, 78.5, and 78.5% of samples to classify correctly (Fig. 2S, Table 4). Using 52 and 131 variables, one L and two M samples were again assigned to group F. Nevertheless, the best classification was obtained using 10 variables in the range 320–341 nm, which probably corresponds to hydroxycinnamic acids. Indeed, Pour Nikfardjam et al. (2003) found that Tokaj varietal wines differ in the content of caftaric, coutaric, and *p*-coumaric acids that are characteristic of this group of compounds. According to other authors, the wavelength regions around 270 and 320 nm were most significant for white wines made from Vilana and Dafni varieties (Philippidis et al. 2017), but wavelengths between 464–490 nm and 670–686 nm were best for Chardonnay, Muscat, and Emir white wines (Sen and Tokatli 2016).

In this part, mono varietal wines F, L, and M were classified. Tokaj selections and other wines are not produced as mono varietal wines. It can be seen from Fig. 1 that the UV–Vis spectra of varietal wines differ very little compared to the variability in the spectra of both Tokaj selections and other wines. This means that production technology has a greater impact than the type of variety. The study of the influence of the vintage year and the aging period in barrels is in progress.

Conclusions

UV–Vis spectrometry combined with chemometry was found to be a useful analytical tool capable of classifying Tokaj wines according to style into essences, Tokaj selections, varietal wines, and other wines, as well as according to the amount of added cibebas into 6-, 5-, 4-, 3-, and 2-putňový, and finally by variety to Furmint, Lipovina, and Muškát žltý. The results demonstrated that the discriminant models developed using selected variables have better prediction ability compared to the models based on the principal components. After selecting the appropriate variables, 100% total correct classification of diluted samples was achieved in all three tasks examined. UV–Vis spectrometry due to its low cost and simplicity can be useful as a screening method for searching for suspicious samples before a more detailed analysis by any of the more sophisticated techniques.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12161-021-02097-y.

Funding This research was supported by the Slovak Research and Development Agency under the contract no. APVV-15–0355. Great thanks to Tokaj wineries Ostrožovič spol. s r.o., Anna Nagyová-Zlatý Strapec and Tokaj & Co., s.r.o. for the cooperation and granting the samples.

Declarations

Ethics Approval This article does not contain any studies with human or animal subjects.

Informed Consent Not applicable.

Conflict of Interest Jana Sádecká declares that she has no conflict of interest. Michaela Jakubíková declares that she has no conflict of interest.

References

Aleixandre-Tudo JL, Nieuwoudt H, Olivieri A, Aleixandre JL, du Toit W (2018) Phenolic profiling of grapes, fermenting samples and wines using UV-Visible spectroscopy with chemometrics. Food Control 85:11–22

Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. Statist Surv 4:40–79

Ballová Ľ, Eftimová Z, Kurhajec S, Eftimová J (2016) Tokaj wines as source of polyphenols with positive effects to health. ТЕОРЕТИЧНІ ТА ПРАКТИЧНІ АСПЕКТИ ДОСЛІДЖЕННЯ

ЛІКАРСЬКИХ РОСЛИН. МАТЕРІАЛИ. II Міжнародної науково-практичної internet-конференції. 21–23 березня 2016. Харків .2016. МІНІСТЕРСТВО ОХОРОНИ ЗДОРОВ'Я УКРАЇНИ. Національний фармацевтичний університет

Bührle F, Gohl A, Weber F (2017) Impact of xanthylium derivatives on the color of white wine. Molecules 22:1376–1393

Carbajal-Ida D, Maury C, Salas E, Siret R, Mehinagic E (2016) Physico-chemical properties of botrytised Chenin blanc grapes to assess the extent of noble rot. Eur Food Res Technol 242:117–126

Chandra S, Chapman J, Power A, Roberts J, Cozzolino D (2017) Origin and regionality of wines—the role of molecular spectroscopy. Food Anal Methods 10:3947–3955

Commission Regulation (EU) No 401/2010 of 7 May 2010 amending and correcting Regulation (EC) No 607/2009 laying down certain detailed rules for the implementation of Council Regulation (EC) No 479/2008 as regards protected designations of origin and geographical indications, traditional terms, labelling and presentation of certain wine sector products. Official J Eur Union OJ L 87, 24(3)2006:2

Campos MP, Sousa R, Pereira AC, Reis MS (2017) Advanced predictive methods for wine age prediction: Part II – A comparison study of multiblock regression approaches. Talanta 171:132–142

Cozzolino D (2015) The role of visible and infrared spectroscopy combined with chemometrics to measure phenolic compounds in grape and wine samples. Molecules 20:726–737

Dankowska A, Kowalewski W (2019) Comparison of different classification methods for analysing fluorescence spectra to characterize type and freshness of olive oils. Eur Food Res Technol 245:745–752

Ferreiro-González M, Ruiz-Rodríguez A, Barbero GF, Ayuso J, Álvarez JA, Palma M, Barroso CG (2019) FT-IR, Vis spectroscopy, color and multivariate analysis for the control of ageing processes in distinctive Spanish wines. Food Chem 277:6–11

Figueiredo-Gonzalez M, Cancho-Grande B, Simal-Gandara J (2013) Effects on colour and phenolic composition of sugar concentration processes in dried-on- or dried-off-vine grapes and their aged or not natural sweet wines. Trends Food Sci Technol 31:36–54

Furdíková K, Machyňáková A, Drtilová T, Špánik I (2020) Comparison of different categories of Slovak Tokaj wines in terms of profiles of volatile organic compounds. Molecules 25:669. https://doi.org/10.3390/molecules25030669

Hu L, Yin Ch, Ma S, Liu Z (2018) Rapid detection of three quality parameters and classification of wine based on Vis-NIR spectroscopy with wavelength selection by ACO and CARS algorithms. Spectrochim Acta A 205:574–581

Ikram ST, Cherukuri AK (2017) Intrusion detection model using fusion of chi-square feature selection and multi class SVM. J King Saud Univ, Comp & Info Sci 29:462–472

Koreňovská M, Suhaj M (2005) Identification of some Slovakian and European wines origin by the use of factor analysis of elemental data. Eur Food Res Technol 221:550–558

Kovács Z, Dinya Z, Antus S (2004) LC–SSI-MS techniques as efficient tools for characterization of nonvolatile phenolic compounds of a special Hungarian wine. J Chromatogr Sci 42:125–129

Law No. 313/2009 on viticulture and wine (Zákon o vinohradníctve a vinárstve) of 30 June 2009 on the web: https://www.aspi.sk/products/lawText/1/69055/1/2

Liu J, Pan TJ, Zhang ZY (2018) Incremental support vector machine combined with ultraviolet-visible spectroscopy for rapid discriminant analysis of red wine. J Spectrosc 2018:1–5

Mazur M, Husáriková L, Kaliňák M, Valko M (2015) "VINUM REGNUM - REX VINORUM" ^{1}H NMR Spectroscopy study of the Slovak Tokaj wines. Czech Chem Soc Symp Ser 13:103–106

Magyar I (2011) Botrytized Wines. Adv Food Nutr Res 63:147–206

Molla N, Bakardzhiyski I, Manolova Y, Bambalov V, Cozzolino D, Antonov L (2017) The effect of path length on the measurement accuracies of wine chemical parameters by UV, visible, and near-infrared spectroscopy. Food Anal Methods 10:1156–1163

Martelo-Vidal MJ, Domínguez-Agis F, Vázquez M (2013) Ultraviolet/visible/near-infrared spectral analysis and chemometric tools for the discrimination of wines between subzones inside a controlled designation of origin: a case study of Rías Baixas. Aust J Grape Wine R 19:62–67

OIV (2016) Compendium of international methods of wine and must analysis. International Organisation of Vine and Wine, Paris

Philippidis A, Poulakis E, Basalekou M, Strataridaki A, Kallithraka S, Velegrakis M (2017) Characterization of Greek wines by ultraviolet-visible absorption spectroscopy and statistical multivariate methods. Anal Lett 50:1950–1963

Pour Nikfardjam MS, László G, Dietrich H (2003) Polyphenols and antioxidative capacity in Hungarian Tokaj wine. Mitt Klosterneubur 53:159–165

Pour Nikfardjam MS, László G, Dietrich H (2006) Resveratrol-derivatives and antioxidative capacity in wines made from botrytized grapes. Food Chem 96:74–79

Rendall R, Pereira AC, Reis MS (2017) Advanced predictive methods for wine age prediction: Part I – A comparison study of single-block regression approaches based on variable selection, penalized regression, latent variables and tree-based ensemble methods. Talanta 171:341–350

Sádecká J, Jakubíková M (2020) Varietal classification of white wines by fluorescence spectroscopy. J Food Sci Technol 57:2545–2553

Sádecká J, Jakubíková M, Májek P (2018) Fluorescence spectroscopy for discrimination of botrytized wines. Food Control 88:75–84

Sen I, Tokatli F (2016) Differentiation of wines with the use of combined data of UV–visible spectra and color characteristics. J Food Compost Anal 45:101–107

Tan J, Li R, Jiang ZT, Zhang Y, Hou YM, Wang YR, Wu X, Gong L (2016) Geographical classification of Chinese Cabernet Sauvignon wines by data fusion of ultraviolet-visible and synchronous fluorescence spectroscopies: the combined use of multiple wavelength differences. Aust J Grape Wine R 22:358–365

Uríčková V, Sádecká J (2015) Determination of geographical origin of alcoholic beverages using ultraviolet, visible and infrared spectroscopy: a review. Spectrochim Acta A 148:131–137

Vabalas A, Gowen E, Poliakoff E, Casson AJ (2019) Machine learning algorithm validation with a limited sample size. PLoS ONE 14(11):e0224365. https://doi.org/10.1371/journal.pone.0224365

Yang J, Sun J, Du L, Chen B, Zhang Z, Shi S, Gong W (2017) Monitoring of paddy rice varieties based on the combination of the laser-induced fluorescence and multivariate analysis. Food Anal Methods 10:2398–2403

Yu J, Wang H, Zhan J, Huang W (2017) Review of recent UV–Vis and infrared spectroscopy researches on wine detection and discrimination. Appl Spectrosc Rev 53:65–86

Zimdars S, Hitschler J, Schieber A, Weber F (2017) Oxidation of wine polyphenols by secretomes of wild *Botrytis cinerea* strains from white and red grape varieties and determination of their specific laccase activity. J Agric Food Chem 65:10582–10590