Journal Pre-proofs

Raman spectroscopy combined with chemometric modeling for predicting physicochemical quality and safety attributes of Asgari grapes

Nazanin Tayyebi1, Fataneh Hashempour-Baltork, Hassan Shojaee-Mend, Zohreh Abdi-Moghadam,

DOI: https://doi.org/ 10.22104/ift.2026.8044.2258

To appear in: Innovative Food Technologies (IFT)

Received Date: 9 December 2025
Revised Date: 5 February 2026
Accepted Date: 12 March 2026

Please cite this article as: Nazanin Tayyebi1, Fataneh Hashempour-Baltork, Hassan Shojaee-Mend, Zohreh Abdi-Moghadam, Raman spectroscopy combined with chemometric modeling for predicting physicochemical quality and safety attributes of Asgari grapes, *Innovative Food Technologies (2026)*, doi: https://doi.org/ 10.22104/ift.2026.8044.2258

Raman spectroscopy combined with chemometric modeling for predicting physicochemical quality and safety attributes of Asgari grapes

Nazanin Tayyebi[1], Fataneh Hashempour-Baltork[2], Hassan Shojaee-Mend [3]*, Zohreh Abdi-Moghadam[1]*

[1] Department of Food Science and Nutrition, Faculty of Medicine, Gonabad University of Medical Sciences, Gonabad, Iran
[2] Halal Research Center of IRI, Iran Food and Drug Administration, Ministry of Health and Medical Education, Tehran, Iran
[3] Department of General Courses, Faculty of Medicine, Gonabad University of Medical Sciences, Gonabad, Iran

*Corresponding authors: Zohreh Abdi-Moghadam; z.abdimoghadam1@gmail.com and Hassan Shojaee-Mend; hassan.shojaee@gmail.com

Abstract

Rapid and low-waste assessment of grape quality remains a challenge for conventional analytical techniques; Raman spectroscopy combined with chemometric modeling provides a promising alternative for predicting the quality attributes of Asgari grapes. Fifty-five grape samples were collected from vineyards across Gonabad, Iran, and key physicochemical parameters—including pH, titratable acidity (TA), flavonoid content, total soluble solids (TSS), anthocyanin content, and the TSS/TA ratio—were measured. Dimensionality reduction was carried out using principal component analysis (PCA) and partial least squares (PLS), followed by the development of predictive models using multiple linear regression (MLR) and support vector machines (SVM). The dataset was randomly divided into training (80%) and testing (20%) subsets. Model performance was evaluated using the coefficient of determination (R^2) and the root mean square error of calibration (RMSEC) and prediction (RMSEP). Among the evaluated models, the PCA–MLR hybrid demonstrated the best predictive performance, achieving R^2 values of 0.75 for anthocyanins, 0.74 for flavonoids, and 0.65 for TSS, with corresponding RMSEC values of 0.56, 0.67, and 1.51, respectively. These results confirm that Raman spectroscopy is a viable and efficient tool for grape quality assessment and represents a practical alternative to conventional chemical analytical methods.

Keywords: multivariate analysis; prediction model; quality control; spectral modeling; Vitis vinifera

1. Introduction

Grapes (Vitis vinifera L.) are among the most extensively cultivated horticultural crops worldwide, occupying more than 6.5 million hectares and producing over 72 million tons annually [1]. In Iran, viticulture covers approximately 113,550 hectares with a production of 1.47 million tons, ranking the country among the leading global producers [1]. Iran is also recognized as a center of grape genetic diversity, hosting numerous local cultivars that are well adapted to

arid and semi-arid climatic conditions [2]. Among these cultivars, the Asgari grape is distinguished by its thin skin, desirable flavor, and suitability for raisin production. It is rich in phytochemicals, including anthocyanins, flavonoids, and organic acids, making it relevant for both food and pharmaceutical applications [3]. Despite its agricultural and economic importance, standardized protocols for quality assessment and optimal harvest timing remain limited for this cultivar.

Traditionally, physicochemical indicators such as total soluble solids (TSS), titratable acidity (TA), and the TSS/TA ratio have been used to evaluate grape maturity and predict consumer acceptability. However, these conventional methods are labor-intensive and time-consuming, which restricts their scalability and limits practical application in precision agriculture systems[4].

In recent years, sensing technologies, particularly vibrational spectroscopy, have emerged as effective tools for the rapid quality assessment of agricultural products. Raman spectroscopy provides a non-invasive technique for analyzing molecular structures through the inelastic scattering of monochromatic light, enabling fast, reagent-free measurements that are suitable for in-field or postharvest monitoring [5]. Its effectiveness in assessing internal quality attributes of fruits has been widely demonstrated. For example, Andersen et al. applied Raman spectroscopy to estimate TSS, TA, and the TSS/TA ratio in strawberries, achieving high predictive accuracy through chemometric modeling. In grapes [6], Gao, et al. [7] developed calibration models for anthocyanin quantification, while Ebrahimi, et al. [8] constructed reliable predictive models for pH and TSS using Raman spectral data. Collectively, these studies highlight the strong potential of Raman spectroscopy for fruit quality evaluation.

From a modeling perspective, the selection of appropriate machine learning algorithms is critical for extracting meaningful relationships between Raman spectral features and grape quality attributes. Raman spectra are high-dimensional and often noisy, and they frequently exhibit non-linear relationships with quality parameters such as anthocyanins and flavonoids[7, 8]. Traditional univariate or simple multivariate methods may fail to capture these subtle patterns, thereby limiting prediction accuracy. Multiple linear regression (MLR) is a widely used chemometric method that is simple, transparent, and robust, particularly when correlated spectral variables are transformed into independent components using principal component analysis (PCA) or partial least squares (PLS) [9]. In contrast, support vector machines (SVM) represent a powerful nonlinear learning algorithm capable of modeling complex and non-linear relationships between spectral data and physicochemical properties, which are commonly observed in biological and agricultural systems [10]. Previous studies have shown that combining Raman spectroscopy with linear models, such as MLR, and nonlinear models, such as SVM, enables a comprehensive evaluation of prediction performance across different quality parameters, including anthocyanins, pH, and soluble solids [7, 8, 11]. Therefore, the simultaneous use of MLR and SVM in the present work allows for a systematic comparison between linear and nonlinear modeling strategies, enhancing the robustness and interpretability of grape quality prediction based on Raman spectral data.

Despite the demonstrated potential of Raman spectroscopy for fruit quality evaluation, existing studies have primarily focused on a limited number of quality attributes, specific grape cultivars, or single predictive models. Moreover, most investigations have addressed either individual chemical compounds or general maturity indicators, rather than providing a comprehensive assessment of multiple quality parameters simultaneously. In addition, there is a lack of studies

focusing on region-specific grape cultivars grown under arid and semi-arid climatic conditions, where environmental variability can significantly influence fruit composition and quality.
Accordingly, the present study aims to evaluate the qualitative attributes of the Asgari grape cultivar grown in Gonabad, Iran, using Raman spectroscopy coupled with advanced multivariate and machine learning models. Unlike previous studies, this work simultaneously predicts multiple key quality parameters, including TSS, pH, anthocyanin and flavonoid content, TA, and the TSS/TA ratio, and compares the performance of different predictive approaches. By integrating a region-specific focus with a comprehensive modeling framework, this study provides a novel and scalable strategy for rapid grape quality assessment, contributing to the advancement of precision agriculture and non-destructive analytical techniques.

2. Materials and Methods

2.1. Sample Preparation

Grapes of the Asgari cultivar (*Vitis vinifera* L.) were collected from vineyards in the Gonabad region, including Ghassabeh, Ghozd, Kakhk, and Mend. A total of 55 samples were collected, with each sample placed in a labeled container. Each sample was washed, dried using absorbent paper, juiced, and filtered. A portion of the juice was reserved for physicochemical analysis, and the remaining portion was allocated for Raman spectroscopy. All samples were stored at 4 °C until testing to minimize postharvest metabolic activity and preserve physicochemical properties, as recommended in table grape postharvest quality maintenance guidelines and studies on postharvest grape handling[12, 13].

2.2. Measurement of physicochemical properties

Total soluble solids (TSS) were measured in triplicate using a digital refractometer (Prismatech, Iran) and reported in degrees Brix (°Brix) according to AOAC Official Method 932.12 [14]. The pH was measured in triplicate using a calibrated pH meter (Consort, Belgium) following AOAC Official Method 981.12 [15]. Titratable acidity (TA) was determined by titration with 0.1 N NaOH according to AOAC Official Method 942.15 [16]. The TSS/TA ratio was calculated as TSS divided by TA [11].

Total flavonoids were quantified using the aluminum chloride colorimetric method following Singleton et al. [17]. Total anthocyanins were measured by the pH differential method according to Giusti & Wrolstad [18].

2.3. Raman measurement

Raman spectra were acquired using a Raman microscope (TDLG100, Teksan, Tehran, Iran) over the spectral range of 4200 to 200 cm^{-1}. A 785 nm laser was used with a power of 300 mW, an integration time of 3 s, and an average of 5 scans to optimize signal intensity. Samples were placed on a sample holder, with the laser focused on the sample. All measurements were performed at room temperature, with a total collection time of 15 s per sample. All samples were scanned under identical conditions.

2.4. Model Preprocessing

Spectral preprocessing comprised baseline correction, smoothing, and normalization.

2.4.1. Baseline Correction

To reduce fluorescence and other baseline artifacts, adaptive iteratively reweighted penalized least squares (airPLS) was applied with parameters optimized for effective baseline removal without over-smoothing peaks. The lambda (λ) value was set to 10^5, the penalty order was 2, and the maximum number of iterations was 10 [19].

2.4.2. Smoothing

After baseline correction, a Savitzky–Golay filter was applied with a window size of 11 points and a polynomial order of 2 to reduce noise while preserving peak shapes [20].

2.4.3. Normalization Using SNV

Spectra were normalized using the standard normal variate (SNV) method. The mean of each spectrum was subtracted, and the resulting values were divided by the corresponding standard deviation to mitigate systematic variations arising from sample preparation and measurement conditions [21].

2.5. Dimensionality Reduction

Given the high dimensionality of the Raman data (838 features per sample), principal component analysis (PCA) and partial least squares (PLS) were used for dimensionality reduction. PCA, an unsupervised method, transformed the data into principal components retaining 95% of the total variance [22]. PLS identified latent variables that maximized covariance between the spectral data and the target variables, including pH, TSS, TA, the TSS/TA ratio, anthocyanin content, and flavonoid content. The optimal number of latent variables was selected by cross-validation [23] .

2.6. Model Development

Two model families were developed using reduced features obtained from PCA and PLS, namely multiple linear regression (MLR) and support vector machines (SVM). Samples were randomly divided into training (80 %) and test (20 %) sets. Given the relatively small sample size (n = 55), random splitting was applied to ensure that the test set remained independent for unbiased performance evaluation. The split was not stratified by vineyard origin or measured physicochemical parameters. Using PCA, five principal components were extracted and used as inputs for both MLR and SVM models. Models were also trained using PLS components, with the number of components optimized for each target variable. MLR was used to establish linear relationships between spectral features and target variables [9].

Model development focused on MLR and SVM due to their complementary strengths. MLR provides interpretable linear relationships between spectral features and response variables, while SVM captures potential nonlinear relationships in the data [9, 10]. Although other regression approaches commonly used in chemometrics, such as partial least squares regression (PLSR) or ensemble methods like random forest, were not evaluated, previous studies have demonstrated that PCA–MLR and SVM models can achieve robust predictive performance for grape quality parameters [7, 8, 24]. Future work may explore additional algorithms to assess potential gains in prediction accuracy or robustness.

SVM hyperparameters were tuned using grid search with cross-validation. The optimal parameters were a regularization parameter (C) of 1 and a kernel coefficient (gamma) of 0.01 [10].

2.7. Model Evaluation

Model performance was assessed using the coefficient of determination (R^2) and the root mean square error (RMSE) [25]. The coefficient of determination indicates the proportion of variance in the observed data explained by the model and was calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \hat{y})^2}$$

where y_i is the observed value, $\hat{y}_i$ is the predicted value, and $\hat{y}$ is the mean of observed values.

RMSE quantifies the average prediction error and was calculated as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

Models were trained on 80 % of the data and evaluated on the remaining 20 % as a holdout test set to ensure unbiased performance assessment. Five-fold cross-validation was additionally applied to select the optimal number of latent variables in the PLS models by minimizing prediction error [25].

3. Results and Discussion

3.1 Physicochemical Analysis

The physicochemical parameters of Asgari grape juice, including pH, total soluble solids (TSS), titratable acidity (TA), the TSS/TA ratio, anthocyanin content, and flavonoid content, are presented in Table 1 as mean values (± SD). Clear regional variations were observed among samples collected from Mend, Ghassabeh, Ghozd, and Kakhk.

The mean pH values followed the order Mend > Ghassabeh > Ghozd > Kakhk, while TSS exhibited a similar trend, ranking Mend > Ghassabeh > Kakhk > Ghozd. In contrast, TA showed an inverse pattern, with the highest values recorded in Kakhk and the lowest in Mend. Consequently, the TSS/TA ratio, which is an important indicator of grape maturity and sensory quality, was highest in Mend and lowest in Kakhk. These trends highlight the combined influence of regional climatic conditions and ripening status on grape physicochemical composition.

The TSS values obtained in this study fall within the ranges reported by Ping, et al. [26], who reported values of 15.47 to 20.28 °Brix, Daniels, et al. [11], who reported values ranging from 10.18 to 24.40 °Brix, and Azadshahraki, et al. [27] , who reported values ranging from 17 to 26.62 °Brix. These results indicate that the soluble solids content of Asgari grapes from Gonabad is comparable to that reported for table grapes grown in other regions. Previous studies have demonstrated that TSS and TA are strongly influenced by temperature and altitude, as increasing temperature and ripeness enhance sugar accumulation while accelerating the degradation of organic acids, thereby reducing TA.

Despite comparable TSS values, TA and the TSS/TA ratio observed in the present study were relatively higher and lower, respectively, than those reported by Ping, et al. [26] , who achieved an R^2 of 0.94 and an RMSEP of 1.96 for TA prediction. A plausible explanation for this difference is the hot and arid climate of the Gonabad region. Under conditions of water stress and elevated temperatures, grapevines tend to accumulate organic acids, particularly tartaric acid, as part of their physiological stress response. This accumulation can result in TSS/TA ratios lower than the commonly cited optimal range of 12.5 to 14 [26, 28]. Similar climatic effects on grape acidity have been reported by Ping, et al. [26]. Moreover, titratable acidity is known to vary dynamically during grape ripening, with tartaric acid levels typically reaching a maximum before gradually declining as ripening progresses. Environmental conditions can modulate this process, leading to region-specific acidity profiles. The observed differences in TA among the studied regions are consistent with the findings of Fernández-Novales, et al. [4], who reported that organic acid composition is strongly influenced by both ripening stage and environmental factors. Grape juice pH, which reflects hydrogen ion concentration and is inversely related to TA, showed values consistent with those reported in the literature.

Table (1)

Grape juice pH reflects hydrogen ion concentration and is inversely related to TA, such that higher acidity corresponds to lower pH [11]. The pH range observed in this study agrees with the values reported by Azadshahraki, et al. [27], who reported a pH range of 3 to 3.96, and Daniels,

et al. [11], who reported a pH range of 3.31 to 4.07, confirming the typical acidity characteristics of table grapes.

Anthocyanin content was highest in samples from Kakhk, which may be attributed to the effect of temperature on anthocyanin stability. Elevated temperatures promote anthocyanin degradation, whereas relatively cooler conditions favor anthocyanin retention. Therefore, the comparatively lower temperatures in Kakhk [29], may have contributed to the higher anthocyanin levels observed. Differences between the present results and those reported by Mirzaei, et al. [30], can also be explained by varietal effects, as their study focused on red seedless grape cultivars, whereas the Asgari cultivar is predominantly green. Flavonoid content was highest in samples collected from Mend. This variation may be associated with enhanced activity of phenylalanine ammonia-lyase [28], a key enzyme in flavonoid biosynthesis, which utilizes phenylalanine as a precursor [28]. a key enzyme in flavonoid biosynthesis that utilizes phenylalanine as a precursor. Environmental and regional factors can influence PAL activity, leading to differences in flavonoid accumulation among grape samples. The lower flavonoid values observed in the present study compared with those reported by Gohari, et al. [31], may be attributed to differences in cultivar type and growing conditions, as grapes contain phenylalanine as a metabolic precursor for PAL and, consequently, for flavonoid biosynthesis.

3.2 Raman spectroscopy analysis

Baseline correction using adaptive iteratively reweighted penalized least squares (airPLS) was applied to a representative spectrum from sample G1 (Fig. 1A). Spectral smoothing was performed using a Savitzky-Golay filter with an 11-point window and a second-order polynomial, achieving effective noise reduction while preserving spectral features (Fig. 1B) [20]. Standard normal variate (SNV) normalization was subsequently applied to center each spectrum at zero mean and unit variance. As illustrated for sample G1 (Fig. 1C), SNV reduced intensity-related variation among samples, thereby facilitating spectral comparison and model development.

Figure (1)

Dimensionality reduction was then performed. Each sample consisted of 838 spectral variables spanning 100 to 3500 cm^{-1}. To mitigate the curse of dimensionality and enhance model robustness, principal component analysis (PCA) and partial least squares (PLS) analysis were employed. PCA retained five principal components explaining 80.2 percent of the total variance (Fig. 2A). For PLS, the optimal number of latent variables for each target parameter was determined using five-fold cross-validation over one to ten components by minimizing the mean squared error (MSE) (Fig. 2B).

Figure (2)

The selected numbers of PLS components were as follows:

1. pH: 1 component
2. TSS: 2 components
3. Flavonoid: 3 components
4. TA: 4 components
5. Anthocyanin: 4 components

These differences reflect varying levels of input-output complexity among the target variables. Final PLS-based models were refitted using the full training dataset with the selected number of components.

3.3 Modeling and Evaluation

Models were trained using 80 percent of the dataset (n = 44) and evaluated on the holdout test set comprising the remaining 20 percent (n = 11). Model performance was assessed using the

coefficient of determination (R^2) and the root mean square error (RMSE). The results obtained for all quality parameters and modeling approaches, including MLR-PCA, SVM-PCA, MLR-PLS, and SVM-PLS, are summarized in Table 2.

Table (2)

For the MLR-PCA model (Fig. 3), the distribution of points around the 1:1 line indicates higher prediction accuracy for flavonoid and anthocyanin content, with larger deviations observed for TA and particularly for pH. In the SVM-PCA model (Fig. 4), pH prediction accuracy improved relative to MLR-PCA, whereas deviations increased for anthocyanin, TA, TSS, and especially flavonoid content. For the MLR-PLS model (Fig. 5), prediction accuracy for TA improved, while anthocyanin, TSS, pH, and particularly flavonoid content exhibited larger errors. In the SVM-PLS model (Fig. 6), most target variables showed wide dispersion around the ideal line, indicating lower predictive accuracy.

.

Figure (3)
Figure (4)
Figure (5)
Figure (6)

Overall, the predictive performance achieved in this study is comparable to, and in some cases slightly exceeds, that reported in previous Raman-based studies of grape quality assessment. Among the evaluated approaches, the MLR-PCA model provided the best performance for anthocyanin (R^2 = 0.75) and flavonoid (R^2 = 0.74) prediction and outperformed SVM-based models for TSS prediction (R^2 = 0.65). The improved performance of the MLR-PCA approach can be attributed to the ability of PCA to denoise spectral data and orthogonalize highly correlated variables, allowing the regression model to focus on the most informative variance components. Similar advantages of PCA-assisted linear modeling have been reported by Radhika, et al. [24] and Gao, et al. [7].

The higher prediction accuracy obtained for anthocyanins and flavonoids, compared with broader maturity indices such as TSS, TA, and pH, highlights a key strength of Raman spectroscopy. As a vibrational spectroscopic technique, Raman spectroscopy provides molecular-level information and is particularly sensitive to specific functional groups, such as O-H and C=O, which are abundant in phenolic compounds. Consequently, quality parameters with distinct chemical fingerprints are predicted more reliably than aggregate indices influenced by multiple biochemical pathways. This observation is consistent with previous studies by Krysa, et al. [32] and Gao, et al. [7] , who also reported superior Raman-based predictions for phenolic compounds relative to general quality parameters.

This behavior can be further explained by the specific Raman band assignments associated with phenolic structures. Raman bands in the region of 3200 to 3400 cm^{-1} are attributed to O-H stretching vibrations of phenolic hydroxyl groups, which are abundant in anthocyanins and flavonoids. Bands located around 1600 to 1650 cm^{-1} correspond to C=C stretching vibrations of aromatic rings, which are characteristic of flavonoid backbones. Additionally, Raman shifts in the range of 1700 to 1750 cm^{-1} are associated with C=O stretching vibrations present in anthocyanin glycosidic structures and related phenolic compounds.

Further diagnostic bands observed near 1300 to 1350 cm^{-1} and 1200 to 1250 cm^{-1} are attributed to C-O stretching and O-H bending modes, which have been reported as key spectral markers for flavonoid and anthocyanin identification in grape-derived matrices. The presence and relative intensity of these bands support the ability of Raman spectroscopy to selectively probe phenolic compounds in aqueous grape juice, despite the dominance of water in the matrix. This spectral-chemical correspondence explains the higher prediction accuracy achieved for anthocyanin and

flavonoid content compared with broader maturity indices such as TSS and TA. These band assignments are consistent with previous FT-Raman and dispersive Raman studies on grape extracts and isolated flavonoids, confirming that the predictive models developed in this study are based on chemically meaningful spectral features rather than purely statistical correlations [33].

For titratable acidity, the MLR-PLS model achieved the best predictive performance ($R^2 = 0.44$), slightly exceeding that of the MLR-PCA ($R^2 = 0.40$) and SVM-PCA ($R^2 = 0.06$) models. This result is consistent with findings reported by Daniels, et al. [11] and Gabrielli, et al. [13], and reflects the strength of PLS in handling collinear predictors and extracting latent variables that are directly relevant to the response variable. In contrast, the SVM-PCA model provided the most accurate prediction for pH ($R^2 = 0.66$), indicating that nonlinear relationships play a more prominent role in pH modeling. Similar conclusions were reported by Ebrahimi, et al. [8] , who demonstrated that combining dimensionality reduction techniques with nonlinear algorithms can enhance predictive accuracy for complex quality attributes.

When compared with near-infrared spectroscopy-based studies on Asgari and other table grape cultivars, the TSS prediction accuracy achieved in this work falls within the range reported in the literature, although it is slightly lower than values obtained in some large-scale NIR studies [11, 26]. This difference can be attributed in part to the relatively small sample size and the inherent complexity of TSS as a quality parameter influenced simultaneously by sugars, acids, and water content. Nevertheless, Raman spectroscopy offers higher chemical selectivity and reduced sensitivity to water, which is particularly advantageous for aqueous matrices such as grape juice. The TSS/TA ratio, as a derived parameter, was not modeled independently, in agreement with Daniels, et al. [11], who emphasized its calculation from directly measured indices as standard practice in grape quality evaluation.

Overall, although the predictive performance of some models does not exceed the highest values reported in large-scale studies, the present work demonstrates a meaningful advancement by simultaneously predicting multiple physicochemical parameters of a region-specific grape cultivar using a unified Raman-chemometric framework. This comprehensive approach, together with the focus on Asgari grapes grown under arid and semi-arid climatic conditions, represents a practical improvement over previous studies that typically addressed single quality attributes or relied on a single modeling strategy.

3.4. Limitations and Future Work

While this study demonstrates the potential of Raman spectroscopy combined with chemometric modeling for predicting key quality attributes of Asgari grapes, several limitations should be acknowledged. The relatively small sample size ($n = 55$) may limit the generalizability and robustness of the developed models. In addition, because the dataset was divided using non-stratified random sampling, vineyard-specific variability may not have been fully balanced between the training and test sets. Future studies employing larger sample sizes should consider stratified or hierarchical sampling strategies to explicitly account for vineyard origin and environmental heterogeneity.

Juice-based measurements were selected to reduce sample heterogeneity and improve spectral reproducibility, and standardized sample preparation and preprocessing procedures were applied to minimize matrix-related variability. Some models, particularly SVM-based approaches for parameters such as TSS and anthocyanin content, exhibited relatively low predictive performance, suggesting potential overfitting or insufficient feature representation.

Despite these limitations, the study provides a solid foundation for future research. Expanding the sample size and including multiple grape cultivars from diverse climatic regions may further improve model accuracy and applicability. Future work could also focus on the development and deployment of portable Raman devices for real-time, in-field quality assessment, enabling non-destructive evaluation of intact berries. The modeling framework presented here may be extended to other fruit crops, supporting broader applications in precision agriculture and postharvest quality control. Integration of Raman spectroscopy with other sensing modalities, such as near-infrared spectroscopy and hyperspectral imaging, may further enhance prediction accuracy and model robustness. These directions hold promise for developing scalable, cost-effective, and rapid tools for quality monitoring in horticultural supply chains.

4. Conclusions

The results of this study indicate that the MLR-PCA model provides the most reliable overall performance for predicting the qualitative attributes of Asgari grapes. This approach demonstrated strong predictive capability for anthocyanin content, flavonoid content, and total soluble solids (TSS), with the highest accuracy observed for anthocyanin prediction ($R^2 = 0.75$).
The SVM-PCA model achieved superior performance in predicting pH, with an R^2 value of 0.66; however, its predictive accuracy for other quality parameters, particularly TSS and titratable acidity (TA), was comparatively lower.
Taken together, these findings confirm that Raman spectroscopy represents a rapid and effective method for assessing the qualitative properties of Asgari grapes. An important advantage of this technique is its suitability for use with portable Raman spectrometers in vineyards, storage facilities, and processing environments, without requiring complex sample preparation. The adoption of this approach can strengthen quality control practices and enhance operational efficiency in agricultural and food-related industries.

Conflict of interest

No conflict of interest has been declared by the authors.

Acknowledgements

The authors thank Gonabad University of Medical Sciences for their financial support of this study, with ethical approval number (IR.GMU.REC.1403.044). Language editing, style harmonization, and clarity improvements were assisted by an AI tool and confirmed by a native English speaker and subject matter expert; the authors remain fully responsible for the scientific content and conclusions.

Author Contributions:

Nazanin Tayebi Data curation, Formal analysis, Methodology, Software, Validation, Writing – original draft; **Fataneh Hashempour-baltork**: Methodology, Project administration, Writing – review and editing; **Hassan Shojaee-Mend**: Data curation, Formal analysis, Methodology, Project administration, Software, Supervision, Writing – review and editing; **Zohreh Abdi Moghaddam**: Project administration, Software, Supervision, Writing – review and editing, Funding acquisition

Funding

This research was funded by the Research and Technology Deputy of Gonabad University of Medical Sciences.

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Ethics declaration

This study was approved by the Ethics Committee of Gonabad University of Medical Sciences under approval number [IR.GMU.REC.1403.044]. All procedures were conducted in accordance with institutional and national ethical standards.

References

[1] FAO. (2023). the state of food security and nutrition in the world 2023. Available: https://www.fao.org

[2] S. M. M. Mirfatah, M. Rasouli, M. Gholami, and A. Mirzakhani, "Phenotypic diversity of some Iranian grape cultivars and genotypes (Vitis vinifera L.) using morpho-phenological, bunch and berry traits " Journal of Horticulture and Postharvest Research, vol. 7, no. Issue 2, pp. 115-140, https://doi.org/10.22077/jhpr.2024.7165.1355 2024. [In Persian]

[3] A. Rahimi, N. Shariatifar, and A. Heshmati, "Fate of pesticide residues during the grape (Vitis vinifera L.) juice production " (in eng), Iranian Journal of Health and Environment, Research vol. 14, no. 3, pp. 413-426, 2021. [In Persian]

[4] J. Fernández-Novales, I. Barrio, and M. P. Diago, "Non-invasive monitoring of berry ripening using on-the-go hyperspectral imaging in the vineyard," Agronomy, vol. 11, no. 12, p. 2534, https://doi.org/10.3390/agronomy11122534 2021.

[5] F. J. Rivero et al., "A Study of Overripe Seed Byproducts from Sun-Dried Grapes by Dispersive Raman Spectroscopy," Foods, vol. 10, no. 3, p. 483, https://doi.org/10.3390/foods10030483 2021.

[6] P. V. Andersen, N. K. Afseth, K. Aaby, M. Ø. Gaarder, S. F. Remberg, and J. P. Wold, "Prediction of chemical and sensory properties in strawberries using Raman spectroscopy," Postharvest Biology and Technology, vol. 201, p. 112370, 10.1016/j.postharvbio.2023.112370 2023/07/01/ 2023.

[7] Z. Gao et al., "Rapid measurement of anthocyanin content in grape and grape Juice: Raman spectroscopy provides Non-destructive, rapid methods," Computers Electronics in Agriculture, vol. 222, p. 109048, https://doi.org/10.1016/j.compag.2024.109048 2024.

[8] I. Ebrahimi, R. de Castro, R. Ehsani, L. Brillante, S. Feng, and F. Research, "Advancing grape chemical analysis through machine learning and multi-sensor spectroscopy," Journal of Agriculture, vol. 16, p. 101085, https://doi.org/10.1016/j.jafr.2024.101085 2024. [In Persian]

[9] N. Draper, Applied regression analysis. McGraw-Hill. Inc, 1998.

[10] V. N. Vapnik, "An overview of statistical learning theory," IEEE transactions on neural networks, vol. 10, no. 5, pp. 988-999, https://doi.org/10.1109/72.788640 1999.

[11] A. J. Daniels, C. Poblete-Echeverría, U. L. Opara, and H. H. Nieuwoudt, "Measuring internal maturity parameters contactless on intact table grape bunches using NIR spectroscopy," Frontiers in Plant Science, vol. 10, p. 1517, https://doi.org/10.3389/fpls.2019.01517 2019.

[12] C. H. Crisosto, J. L. Smilanick, D. A. Luvisi, and N. K. Dokoozlian, "Maintaining table grape post-harvest quality for long distant markets," in Proceedings of the International Symposium on Table Grape Production: 1994 june 28 & 29, Anaheim, California, 1994, pp. 195-199: American Society for Enology and Viticulture, ASEV.

[13] M. Gabrielli, V. Lançon-Verdier, P. Picouet, and C. Maury, "Hyperspectral imaging to characterize table grapes," Chemosensors, vol. 9, no. 4, p. 71, https://doi.org/10.3390/chemosensors9040071 2021.

[14] AOAC, Official Method 932.12, "Refractometric determination of soluble solids in fruit juices". Gaithersburg, MD, USA AOAC International, 2019.

[15] AOAC, Official Method 981.12, "Measurement of pH in fruit juices,". Gaithersburg, MD, USA AOAC International, 2019.

[16] AOAC, Official Method 942.15, "Acidimetric determination of titratable acidity,". Gaithersburg, MD, USA: AOAC International, 2019.

[17] V. L. Singleton, R. Orthofer, and R. M. Lamuela-Raventós, "[14] Analysis of total phenols and other oxidation substrates and antioxidants by means of folin-ciocalteu reagent," in Methods in enzymology, vol. 299: Elsevier, 1999, pp. 152-178.
[18] M. M. Giusti and R. E. Wrolstad, "Characterization and measurement of anthocyanins by UV-visible spectroscopy," Current protocols in food analytical chemistry, no. 1, pp. F1. 2.1-F1. 2.13, https://doi.org/10.1002/0471142913.faf0102s00 2001.
[19] Z.-M. Zhang, S. Chen, and Y.-Z. Liang, "Baseline correction using adaptive iteratively reweighted penalized least squares," Analyst, 10.1039/B922045C vol. 135, no. 5, pp. 1138-1146, https://doi.org/10.1039/B922045C 2010.
[20] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures," Analytical chemistry, vol. 36, no. 8, pp. 1627-1639, https://doi.org/10.1021/ac60319a045 1964.
[21] R. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra," Applied spectroscopy, vol. 43, no. 5, pp. 772-777, https://doi.org/10.1366/0003702894202201 1989.
[22] I. T. Jolliffe, Principal component analysis for special types of data. Springer, 2002.
[23] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," Chemometrics intelligent laboratory systems, vol. 58, no. 2, pp. 109-130, https://doi.org/10.1016/S0169-7439(01)00155-1 2001.
[24] A. Radhika, M. S. Masood, and Applications, "Effective dimensionality reduction by using soft computing method in data mining techniques," Soft Computing-A Fusion of Foundations, Methodologies, vol. 25, no. 6, https://doi.org/10.1007/s00500-024-10232-0 2021.
[25] V. A. Barbur, "Introduction to Linear Regression Analysis," ed: JSTOR, 1994.
[26] F. Ping et al., "Quality Assessment and Ripeness Prediction of Table Grapes Using Visible–Near-Infrared Spectroscopy," Foods, vol. 12, no. 12, p. 2364, https://doi.org/10.3390/foods12122364 2023.
[27] F. Azadshahraki, S. Kalantari, B. Jamshidi, and S. Research, "Nondestructive Quality Assessment of Grape (Cv. Asgari) Based on NIR Spectroscopy," Agricultural Mechanization, vol. 19, no. 70, pp. 13-24, https://doi.org/10.22092/erams.2017.105874.1080 2018. [In Persian]
[28] S. Poni et al., "Grapevine quality: A multiple choice issue," Scientia horticulturae, vol. 234, pp. 445-462, https://doi.org/10.1016/j.scienta.2017.12.035 2018.
[29] M. Ajam, "Geography , tribes and minorities in khorasan," university of Imam Sadegh,Tehran, https://doi.org/10.1016/j.postharvbio.2023.112370 1992. [In Persian]
[30] F. Mirzaei, R. Karimi, M. Rasouli, M. Rostami, and M. Koulivand, "Comparision of Total Soluble Solid, Total Phenols, Flavonoids and Anthocyanins of Five Grapevine Cultivar In Malayer Vineyards," 2015. [In Persian]
[31] G. Gohari, S. Freydoni, S. Panahirad, N. Sepehri, and M. R. Dadpour, "Foliar application of phenylalanine on nutritional value in Vitis vinifera var. Hosseini," Food Research Journal, vol. 30, no. 4, pp. 109-121, https://doi.org/10.22034/fr.2021.37105.1708 2021.
[32] M. Krysa, M. Szymańska-Chargot, and A. Zdunek, "FT-IR and FT-Raman fingerprints of flavonoids–a review," Food chemistry, vol. 393, p. 133430, https://doi.org/10.1016/j.foodchem.2022.133430 2022.
[33] J. Nogales-Bueno, B. Baca-Bocanegra, A. Rooney, J. M. Hernández-Hierro, H. J. Byrne, and F. J. Heredia, "Study of phenolic extractability in grape seeds by means of ATR-FTIR and Raman spectroscopy," Food Chemistry, vol. 232, pp. 602-609, https://doi.org/10.1016/j.foodchem.2017.04.049 2017.

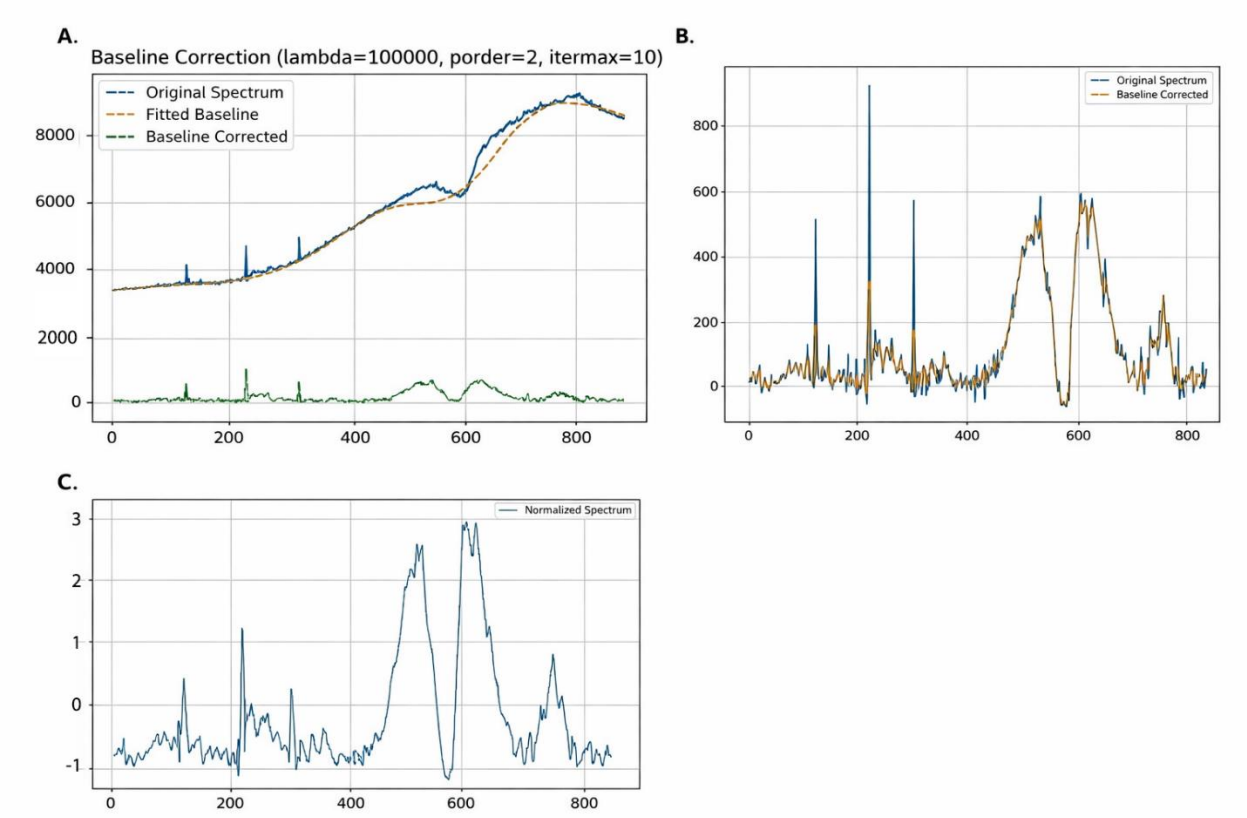


Figure 1- Preprocessing stages applied to Raman spectroscopy data of grape sample G1. (A) Spectrum after baseline correction using the airPLS method. (B) Smoothed spectrum using a Savitzky–Golay filter. (C) Effect of Standard Normal Variate (SNV) normalization on the spectral values.

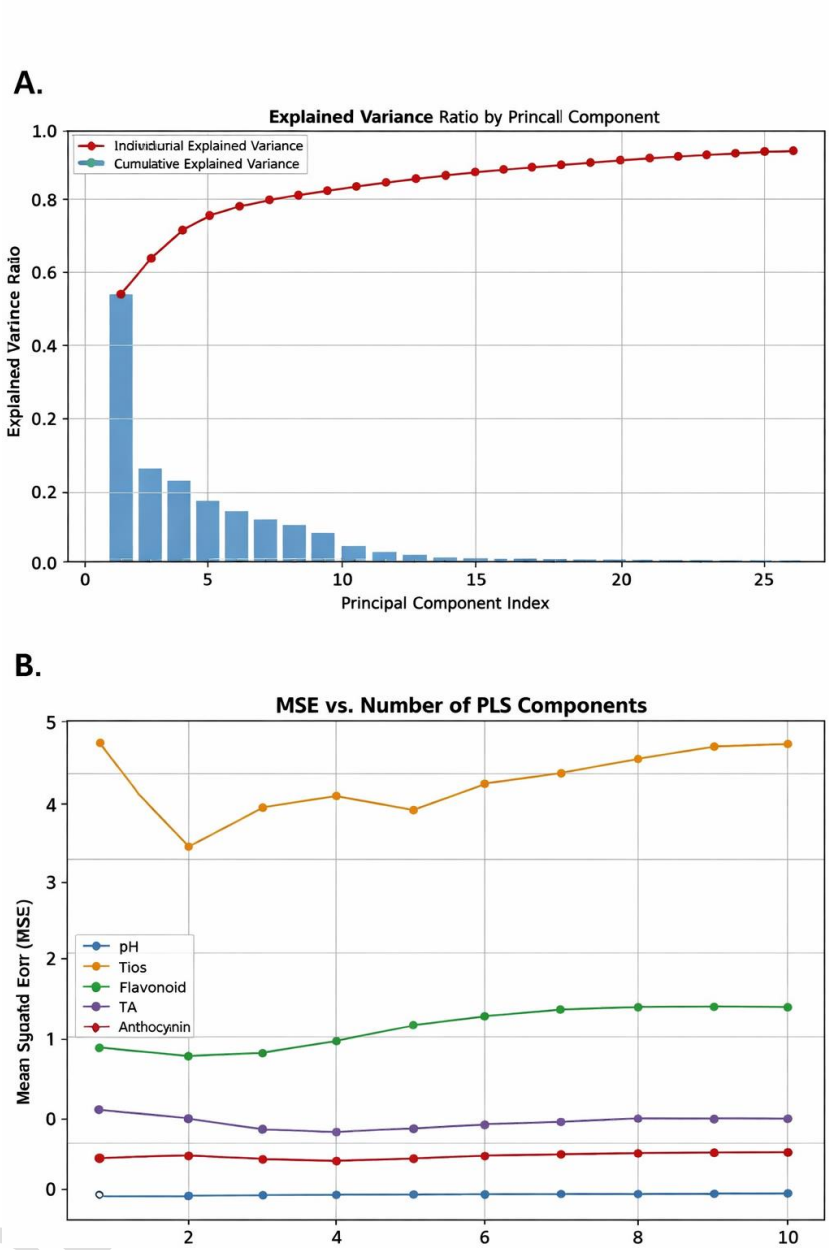


Figure 2 - Dimensionality reduction of Raman spectral data. (A) Variance explained by each principal component. (B) Cross-validation results for selecting optimal PLS components per output variable.

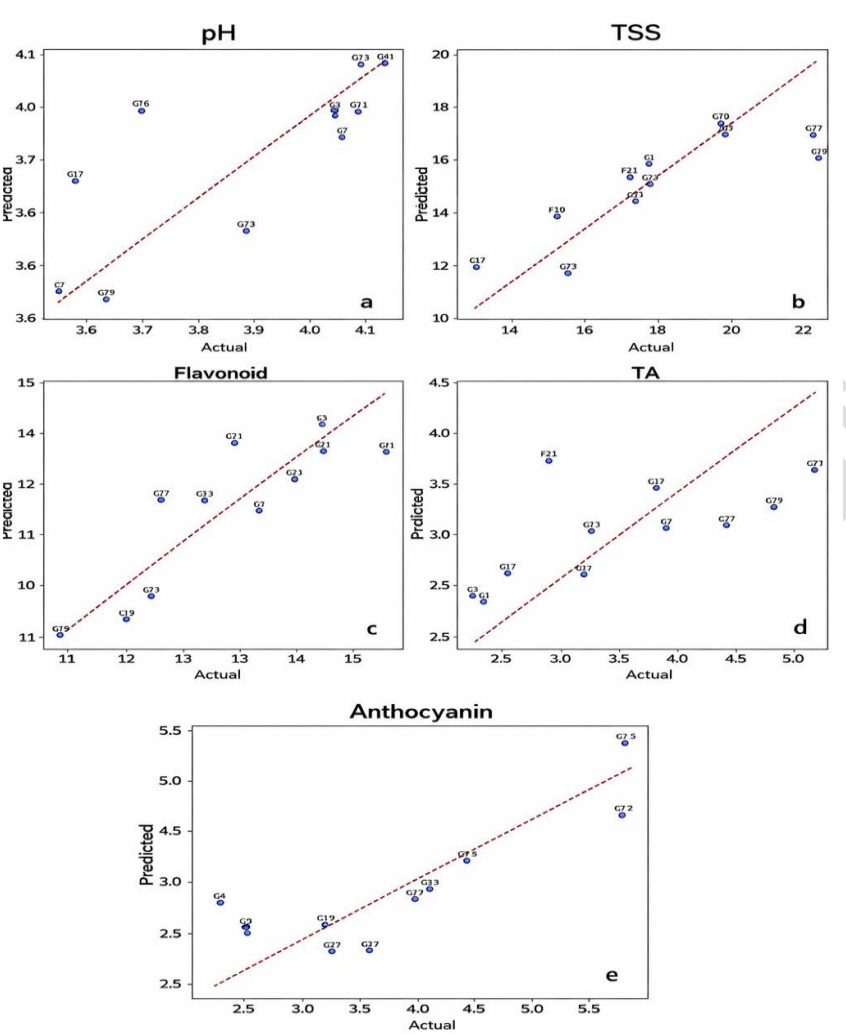
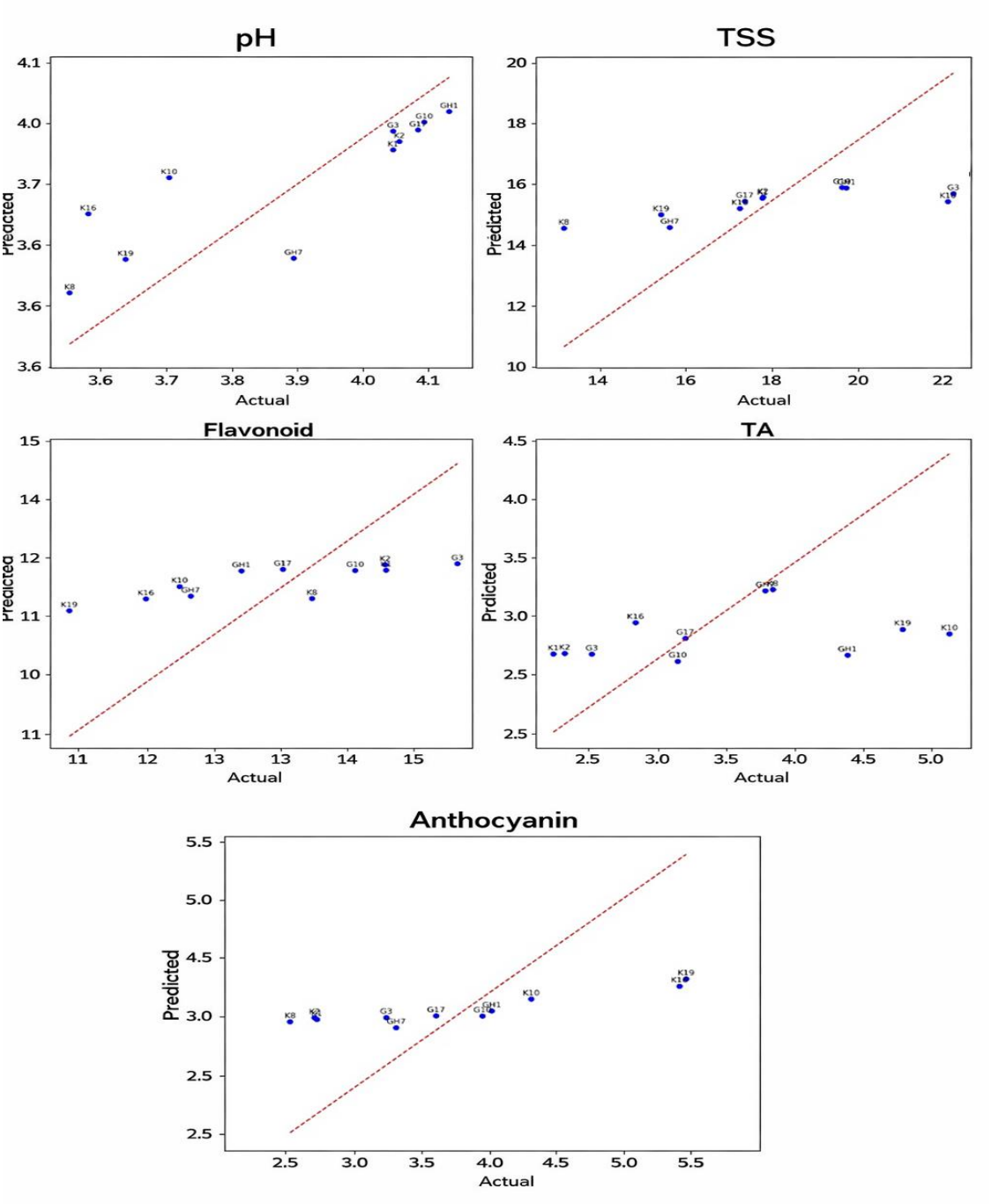


Figure 3 - Prediction of quality parameters in 11 grape samples from the test set using the MLR-PCA model. (a) pH, (b) total soluble solids (TSS), (c) flavonoids, (d) titratable acidity (TA), (e) anthocyanins.



Prediction of quality parameters in 11 grape samples from the test set using the SVM-PCA model. (a) Figure 4 - **pH, (b) total soluble solids (TSS), (c) flavonoids, (d) titratable acidity (TA), (e) anthocyanins.**

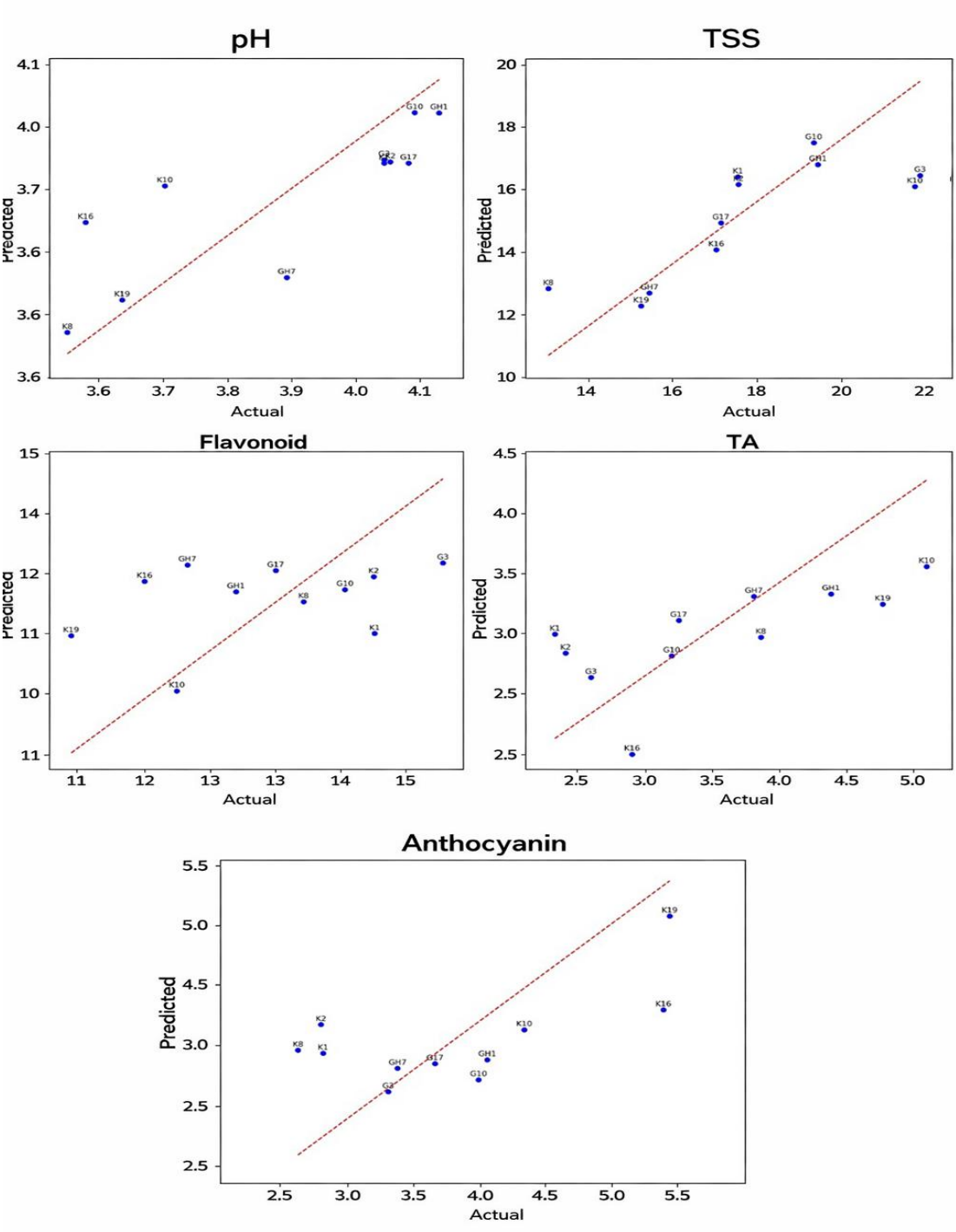


Figure 5 - **Prediction of quality parameters in 11 grape samples from the test set using the MLR-PLS model. (a) pH, (b) total soluble solids (TSS), (c) flavonoids, (d) titratable acidity (TA), (e) anthocyanins.**

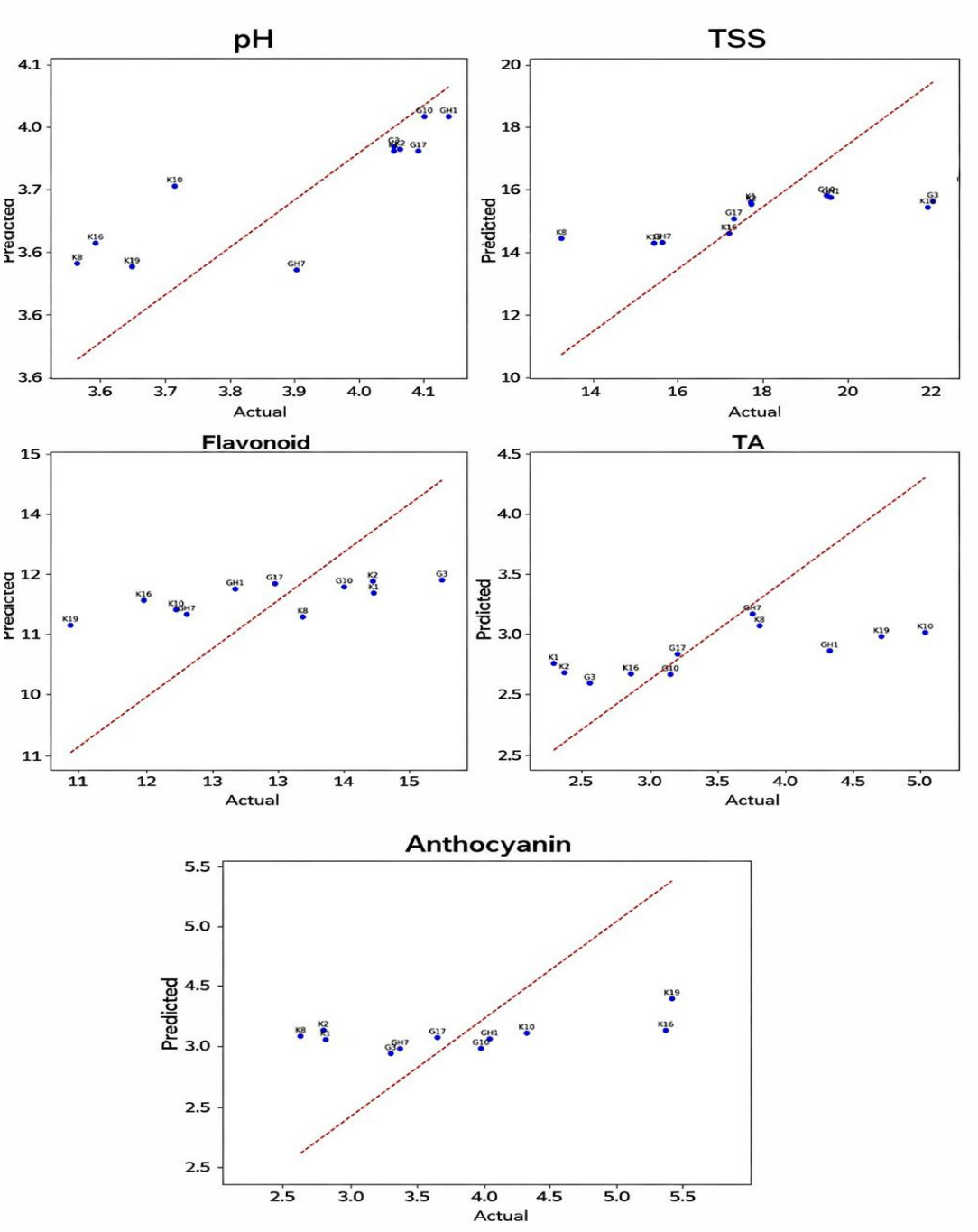


Figure 6 - **Prediction of quality parameters in 11 grape samples from the test set using the SVM-PLS model. (a) pH, (b) total soluble solids (TSS), (c) flavonoids, (d) titratable acidity (TA), (e) anthocyanins.**

Table 1
physicochemical characteristics of Asgari grape samples

Vineyard Name*	TA (100 ml/g)	Deviation of pH	TSS (°Brix)	TSS/TA Ratio	Anthocyanin (mg/l)	Flavonoid (µg/ml)
Ghozd (G)	3.1694 ± 0.5293**	3.9299 ± 0.31649	16.6082 ± 2.6518	5.4829 ± 1.5940	3.6341 ± 0.8658	14.2917 ± 0.4081
Ghassabeh (Gh)	3.1637± 0.8824	4.0275 ± 0.20183	17.6887 ± 2.0882	6.0262 ± 1.9386	3.1287 ± 0.6308	13.3412 ± 0.2948
Kakhk (K)	3.9095±0.75630	3.7938 ± 0.1916	17.5861 ± 2.4157	4.6361 ± 0.9476	4.4038 ± 1.3580	11.7380 ± 1.3002
Mend (M)	2.4811 ± 0.1765	4.1522 ± 0.12715	18.6933 ± 3.1168	7.4988 ± 0.8757	3.2188 ± 0.2074	15.3866 ± 0.4990

*In this study, 17 samples were collected from Ghozd (G1_G17), 8 from Ghassabeh (Gh1_Gh8), 21 from Kakhk (K1_K21), and 9 from Mend (M1_M9)
** Mean ± Standard Deviation

Table 2. Evaluation results of all models for all the factors examined in this study

Facator	MLR-PCA		MLR-PLS		SVM-PLS		SVM-PCA	
	R2	**RMSE**	**R2**	**RMSE**	**R2**	**RMSE**	**R2**	**RMSE**
pH	0.58	0.15	0.64	0.14	0.63	0.14	0.66	0.13
TSS	0.65	1.51	0.61	1.60	0.30	2.16	0.23	2.26
Flavonoid	0.74	0.67	0.11	1.23	0.23	1.15	0.27	1.12
TA	0.40	0.67	0.44	0.65	0.25	0.75	0.06	0.84
Anthocyanin	0.75	0.56	0.39	0.88	0.15	1.03	0.26	0.96