EXTENDED ABSTRACT

Advancing wine authentication: non-invasive near-infrared spectroscopy and machine learning for vintage and quality traits assessment

Claudia Gonzalez Viejo[1,2,*], Natalie Harris[1], Alexis Pang[3], and Sigfredo Fuentes[1,2,4]
*Corresponding author: cgonzalez2@unimelb.edu.au

[1] Digital Agriculture Food and Wine Group, School of Agriculture, Food and Ecosystem Sciences, Faculty of Science. The University of Melbourne, Parkville, VIC 3010, Australia
[2] Centre of Excellence in Plants for Space. Australian Research Council, The University of Melbourne Node. The University of Adelaide (Lead University), Glen Osmond Rd, Adelaide, SA, Australia
[3] School of Agriculture, Food and Ecosystem Sciences, Faculty of Science. The University of Melbourne, Parkville, VIC 3010, Australia
[4] Tecnologico de Monterrey, School of Engineering and Science, Ave. Eugenio Garza Sada 2501, Monterrey, N.L., Mexico

Keywords: wine fraud, digital technologies, Shiraz, rapid methods, provenance

INTRODUCTION

The possibility of fraud and adulteration poses a significant risk within the global food industry, with wine ranking among the top counterfeited or adulterated products within the food and beverage industries. It is estimated that around 20% of wines currently on the market may be fraudulent, resulting in economic losses exceeding $65 billion by 2024 (Liska 2022, Gopalakrishna Pillai, Ngcobo-Onunkwo et al. 2024). Beyond the financial impact, wine fraud undermines consumer trust and could damage a brand's reputation.

Different types of fraud are found in the food industry, with the most common in the wine sector being adulteration and counterfeiting. Adulteration typically involves the addition of unauthorized ingredients and/or additives such as sugars, water (dilution), and pigments, which alter their physicochemical parameters and sensory descriptors. On the other hand, counterfeiting is usually related to provenance, which involves wine origin, grape variety, and vintage, and marketing low-quality or low-cost wine as expensive premium products (Gopalakrishna Pillai, Ngcobo-Onunkwo et al. 2024).

Various analytical approaches have been developed to address these challenges of detecting wine adulteration and counterfeiting. Traditional techniques include physicochemical analysis, isotopic ratio determination, inductively coupled plasma mass spectrometry (ICP-MS), and gas or liquid chromatography. More advanced identification at the molecular level has also been used through DNA-based assays to verify grape variety and origin. Furthermore, some spectroscopic techniques include infrared spectroscopy, nuclear magnetic resonance, atomic spectroscopy, and near infrared (NIR) spectroscopy, which have gained more interest due to their reliability. On the other hand, emerging technologies offering rapid, often low-cost alternatives are sensor-based, such as electronic noses and tongues. Most of these, especially the spectroscopic and sensor-based techniques, are usually coupled with machine learning modelling such as different variations of partial least squares (PLS), linear discriminant analysis (LDA), and artificial neural networks (ANN) (Sun, Zhang et al. 2022, Mac, Pham et al. 2023).

A disadvantage of most existing techniques is that they are invasive, meaning that they require opening the bottles for the assessment, limiting the analysis to just a few bottles. To account for this, a non-invasive method using ultraviolet-visible (UV-Vis), and NIR spectroscopy through the bottle has been presented for assessing oxidation in white wine within 400 – 1100 nm and coupled with principal components analysis to group samples (Cozzolino, Kwiatkowski et al. 2007). More recently, an improved approach using NIR within (1596 – 2396 nm) through the bottle coupled with artificial neural networks has been proposed by our research group to predict vintage, sensory descriptors, and volatile aromatic compounds in red wine (Harris, Gonzalez Viejo et al. 2023), and beer (Harris, Gonzalez Viejo et al. 2025).

RESEARCH OBJECTIVES

This study aimed to develop a novel non-invasive method using a portable handheld NIR spectroscopy device to analyze Shiraz wines through the bottle for quality and vintage assessment. Coupled with machine learning modelling, the NIR absorbance values of wines from the same Australian winery were used to predict the wine vintage (Model 1) and the intensity of sensory descriptors (Model 2) as quality traits.

MATERIAL AND METHODS

Samples description

Samples from a total of 13 vintages (2000 – 2021) of Shiraz wine from the Dookie College Winery, Victoria, Australia (The University of Melbourne, 36° 38′S, 145° 71′E) were used for this study. Vintages were from 2000, 2007, 2008, 2010, and 2013 – 2021 and contained within 13.8 and 14.9% alcohol. All vintages were produced using the same techniques and conditions, and a standard amber Bordeaux bottle of 750 mL was used.

Near-infrared spectroscopy

A NIR spectroscopy handheld device (MicroPhazir RX analyzer, Thermo Fisher Scientific, Waltham, MA, USA) with a custom-made attachment to avoid external light was used to measure the chemical fingerprinting of the wine through the bottle. To account for the glass effects, the empty bottles were measured, and these values were subtracted from the wine through bottle values. For calibration purposes, a spectralon was used before the first sample and after 10 – 15 measurements. All samples were measured at three points of the bottle (top, middle, and bottom) and three measurements per position (n = 27).

Sensory analysis

A sensory session was conducted with 12 trained panelists under the quantitative descriptive analysis (QDA®) method. All participants were form The University of Melbourne (UoM) and were asked to sign a consent form (Ethics ID: 1953926.4). The sensory questionnaire was displayed in tablets using the BioSensory© application (app; The University of Melbourne, Parkville, VIC, Australia) and consisted of rating the intensity of 20 sensory descriptors using a 15-cm non-structured scale. These descriptors involved visual (clarity, color intensity), aromas (truffle, smoke, blackberry, blackcurrant, prune, butter, pepper, cedar, violet, redcurrant), basic tastes (bitter, acidic, sweet), mouthfeel (astringency, body, warming, tingling), and perceived quality.

Machine learning modelling

Two machine learning models based on artificial neural networks (ANN) were developed using the NIR absorbance values as inputs to predict the (i) wine vintages (Model 1) using a classification model and (ii) intensity of sensory descriptors (Model 2) using a regression model.

Models were developed using a code written in Matlab® R2021a (Mathworks, Inc., Natick, MA, USA) by UoM's Digital Agriculture Food and Wine research group. Model 1 was constructed using the Bayesian Regularization algorithm with interleaved data division, with 70% of samples used for training and 30% for testing. Model 2 was developed using the Levenberg-Marquardt algorithm with random data division and 70% of samples used for training, 15% for validation, and 15% for testing. A neuron trimming was done for both models using 3 to 10 neurons to find the best model with no under- and overfitting.

RESULTS

Table 1 shows that Model 1 was highly accurate (97%) at predicting the vintage of the Shiraz wines. It showed no signs of under- or overfitting since the performance measured as a mean squared error (MSE) value of the training stage was lower (MSE < 0.001) than the testing stage (MSE = 0.008). In the receiver operating characteristics (ROC) shown in Figure 1a, it can be observed that most vintages are at the true positive rate, with vintage 2016 being the highest for false positive rate.

Model 2 also presented a very high correlation coefficient (R = 0.95) for predicting the intensity of 20 sensory descriptors. As Table 1 shows, the slope (b) at all stages of the model (training, validation, testing, and overall) was high and close to one (b = 0.89 – 0.96). Furthermore, the MSE value for the training stage (MSE = 0.09) was lower than the validation (MSE = 0.60) and training (MSE = 0.53) stages. The latter two are close to each other, confirming there was no under- or overfitting. Figure 1b shows the overall model with the 95% predictive bounds. As can be seen, the model presented 5% of outliers (351 out of 7020 data points), with color and clarity as the highest in outliers. This was expected since the NIR spectra range used was within 1596 – 2396 nm, and those color-related parameters are usually found in the UV-Vis spectra; however, it was still able to predict these descriptors accurately.

CONCLUSION

The novel method presented in this study offers a cost-effective and rapid solution for winemakers and retailers to evaluate wine quality and authenticity without the need to open bottles. Furthermore, it holds potential for expanding assessments to include additional authentication criteria such as region, country of origin, and grape variety, thus enhancing the integrity of the wine market.

REFERENCES

Cozzolino, D., et al. (2007). "A feasibility study on the use of visible and short wavelengths in the near-infrared region for the non-destructive measurement of wine composition." Analytical and bioanalytical chemistry **387**(6): 2289-2295.

Gopalakrishna Pillai, S., et al. (2024). "Transparency Uncorked: Leveraging Blockchain to Tackle International Wine Fraud." Journal of Hospitality & Tourism Cases: 21649987241290984.

Harris, N., et al. (2023). "Non-Invasive Digital Technologies to Assess Wine Quality Traits and Provenance through the Bottle." Fermentation **9**(1): 10.

Harris, N., et al. (2025). "Enhancing beer authentication, quality, and control assessment using non-invasive spectroscopy through bottle and machine learning modeling." Journal of food science **90**(1): e17670.

Liska, A. (2022). "Counterfeit Wine, Spirits, and Cheese." Retrieved 30 January 2023, 2023, from https://www.recordedfuture.com/counterfeit-wine-spirits-cheese.

Mac, H. X., et al. (2023). "Current techniques for fruit juice and wine adulterant detection and authentication." Beverages **9**(4): 84.

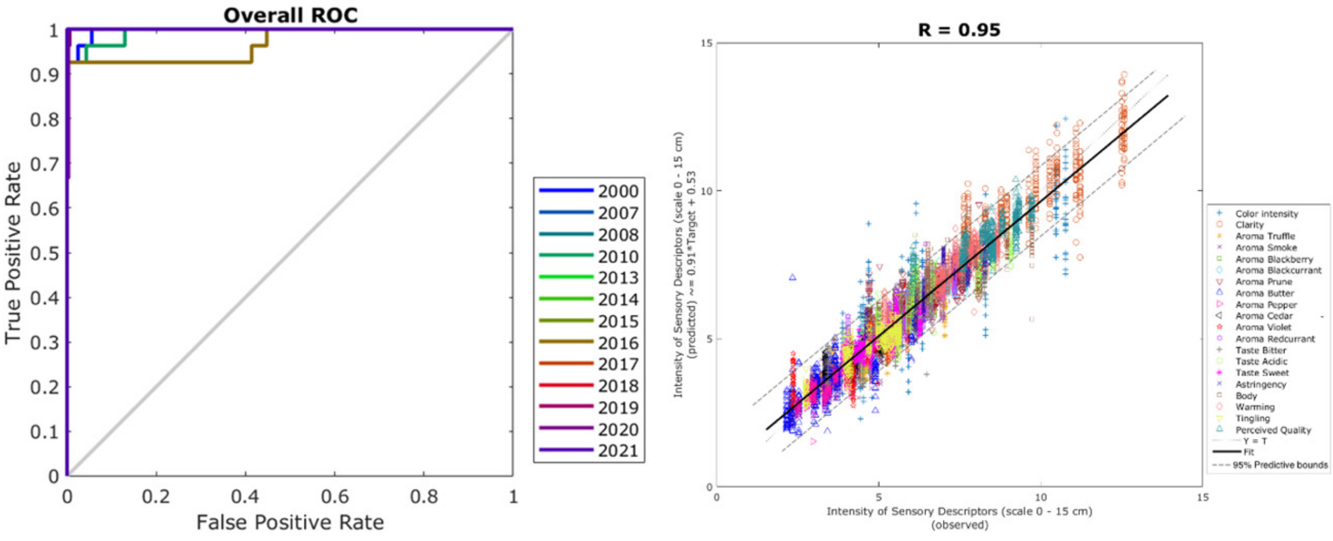
Sun, X., et al. (2022). "Real wine or not? Protecting wine with traceability and authenticity for consumers: chemical and technical basis, technique applications, challenge, and perspectives." Critical reviews in food science and nutrition **62**(24): 6783-6808.

TABLES AND FIGURES

Table 1. Statistical results of the artificial neural network models showing the accuracy and performance of each stage. Abbreviations: MSE: mean squared error; R: correlation coefficient.

Model 1: Vintages				
Stage	**Samples**	**Accuracy**	**Error**	**Performance (MSE)**
Training	246	99.2%	0.8%	<0.001
Testing	105	92.4%	7.6%	0.008
Overall	351	97.1%	2.8%	-

Model 2: Sensory Descriptors					
Stage	**Samples**	**Observations**	**R**	**Slope**	**Performance MSE**
Training	245	4900	0.96	0.92	0.09
Validation	53	1060	0.92	0.91	0.60
Testing	53	1060	0.92	0.89	0.53
Overall	351	7020	0.95	0.91	-



(a) (b)

Figure 1. Graphical results of the (a) Model 1 showing the receiver operating characteristics (ROC) curve for vintage prediction, and (b) Model 2 showing the regression figure with each sensory descriptor represented by a different color and symbol. Abbreviations: R: correlation coefficient.