 https://doi.org/10.1051/bioconf/20235602003

Near-infrared spectroscopy analysis of wines through bottles to assess quality traits and provenance

Claudia Gonzalez Viejo, Natalie Harris, and Sigfredo Fuentes

Digital Agriculture, Food and Wine Sciences Group. School of Agriculture and Food, Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Melbourne, VIC 3010, Australia

Abstract. Due to increased fraud rates through counterfeiting and adulteration of quality wines, it is important to develop novel non-destructive techniques to assess wine quality and provenance. Therefore, our research group developed a novel method using near-infrared (NIR) spectroscopy (1596-2396 nm) coupled with machine learning (ML) modeling to assess wine vintages and quality traits based on the intensity of sensory descriptors through the bottle. These were developed using samples from an Australian vineyard for Shirazwines. Models resulted in high accuracy 97% for classification (vintages) and R=0.95 regression (sensory quality traits). The proposed method will allow to assess authenticity and sensory quality traits of any wines in the market without the need to open the bottles, which is rapid, accurate, effective, and convenient. Furthermore, currently, there are low-cost NIR devices available in the market with the required spectral range and sensitivity, which can be affordable for winemakers and retailers that can be used with the ML models proposed here.

1 Introduction

There are high complexities in the attempt to define wine quality. However, physical-chemometric techniques can objectively assess wine quality traits [1]. Sensory analysis using trained panels could also broaden the definition of wines' specific characteristics that may help define quality traits.

These techniques have also been implemented to assess the provenance and counterfeiting of wines or to detect mislabeling from wineries. Near-infrared spectroscopy can offer a complete chemical fingerprinting profile of wines depending on the spectral range and sensitivity of different instruments. This instrument has been used for various applications, from general quality assessment [2] to predicting aroma profiles in beverages [3]. Furthermore, the use of digital technologies and sensors, such as NIR, electronic noses, and tongues, allows the implementation of Artificial Intelligence (AI) and ML modeling for wine quality prediction [4], to the specific prediction of phenolic compounds in red wines and geographical provenance [5], sugar content [5] astringency [6], and wine authenticity [7].

However, most of these works are based on wine samples extracted by direct measurement from wines or by opening wine bottles for measurement purposes. There is limited research focused on measurements through the bottle that can offer non-invasive tools for physicochemical, sensory, and aroma profile assessment of wines. This work presents the implementation of NIR spectroscopy measured through the bottles of unopened wines in parallel with AI and ML techniques to assess wine quality traits, and provenance and as a potential method to evaluate counterfeiting and mislabeling.

2 Materials and methods

2.1 Samples description

Samples of 13 different vintages from 2000-2021 produced at the Dookie College Winery from The University of Melbourne, Victoria, Australia (36°38′S, 145°71′E) were used for this study.

2.2 Near-infrared spectroscopy and sensory analysis

Samples were measured to assess their chemical fingerprinting within 1596-2396 through the bottle at three different spots (top, middle, and bottom) and three different sides (three measurements per spot/side; n = 27) using a handheld NIR Micro PHAZIR™ RX Analyzer (Thermo Fisher Scientific, Waltham, MA, USA) with a custom-made attachment to cover any external light and reduce noise from the environment. The empty bottles were measured three times and averaged further to subtract the values from the wine through bottle data to remove the glass components.

A quantitative descriptive analysis (QDA®) sensory session was conducted with a trained panel of 12 participants. The study was approved by the Human Ethics Advisory Group from The University of Melbourne (ID: 1953926.4), and panelists signed a consent form to participate. A total of 20 descriptors, (i) color intensity, (ii) clarity, (iii) aroma truffle, (iv) aroma smoke, (v) aroma blackberry, (vi) aroma blackcurrant, (vii) aroma prune, (viii) aroma butter, (ix) aroma pepper, (x) aroma cedar, (xi) aroma violet, (xii) aroma redcurrant, (xiii) bitterness, (xiv) acidic, (xv) sweetness, (xvi) astringency, (xvii) body, (xviii) warming mouthfeel, (xix) tingling mouthfeel, and

https://doi.org/10.1051/bioconf/20235602003

(xx) perceived quality, were evaluated using a 15-cm non-structured scale.

2.3 Machine learning modelling

Two machine learning models were developed based on artificial neural networks (ANN) using a code written in Matlab® R2021a (Mathworks, Inc., Natick, MA, USA) by the Digital Agriculture, Food and Wine group from The University of Melbourne (DAFW-UoM) [8]. Model 1 was developed using pattern recognition with Bayesian Regularization algorithm and NIR data as inputs to predict the vintages (Fig. 1a). Model 2 was developed using regression ANN with the Levenberg Marquardt algorithm and NIR data as inputs to predict the intensity of 20 sensory descriptors (Fig. 1b). Both models were developed with interleaved data division as 70% for training and 30% for testing in Model 1 and random data division as 70% training, 15% validation, and 15% testing for Model 2. Model 1 performance was assessed using cross-entropy, while Model 2 used a means squared error algorithm.

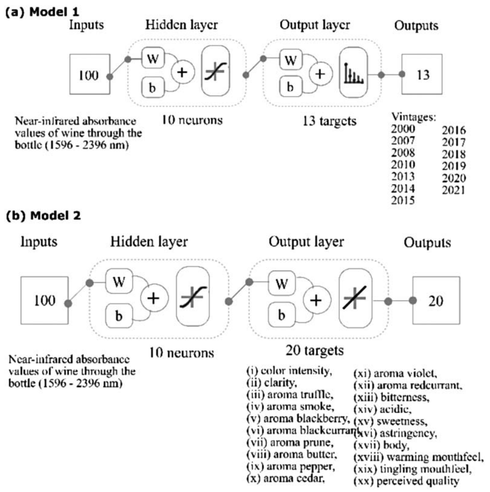

Figure 1. Diagrams of the artificial neural network (a) classification Model 1 to predict the wine vintages and (b) regression Model 2 to predict the intensity of 20 sensory descriptors using the near-infrared absorbance values measured through the bottle.

3 Results and Discussion

Table 1 shows that Model 1 had a very high accuracy (97%) in predicting the vintages. This model did not present signs of over-fitting since the performance (cross-entropy) value of the training stage (<0.01) was lower than the testing (0.01).

Table 1. Results from the classification machine learning model to predict wine vintages.

Stage	Samples	Accuracy	Error	Cross-Entropy
Training	246	99.2%	0.8%	<0.01
Testing	105	92.4%	7.6%	0.01
Overall	351	97.2%	2.8%	-

The ML models obtained, due to their accuracy and without signs of over fitting, are highly replicable since bottle material and thickness are usually consistent with the commercial source by the winery used in this study. Models would need to be retrained for wineries using different shapes of bottles or internal and external diameters. These ML models would not apply to sparkling wines since bottles are thicker to sustain pressures of around 8 bars or more. The latter application would require retraining the algorithm incorporating more data using these types of bottles.

Figure 2 shows the receiver operating characteristics curve where all vintages were close to the true-positive rates. This confirms the model accuracy and suitability to classify the samples according to their vintage.

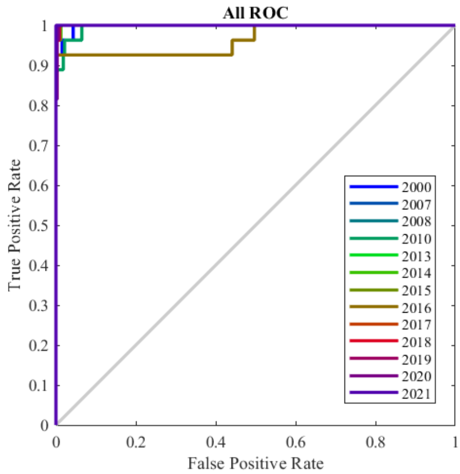

Figure 2. Overall receiver operating characteristics (ROC) curve showing the true-positive and false-positive rates of Model 1 for all vintages.

It is well known that variability of weather conditions and water availability within seasonsdirectly impact the quality traits of grapes produced and the subsequent winemaking process. Seasonal variability made it possible for the ML models to establish a pattern of analysis to distinguish different vintages from the same winery, which can be used to establish consistency and traceability of different bottles to avoid mislabeling.

Table 2 shows that Model 2 had accuracy based on the correlation coefficient (R = 0.95; Fig. 3) to predict the intensity of 20 sensory descriptors. It can be observed that the slopes of all stages are high (>0.88), and the performance value of the testing stage (MSE = 0.26) was lower than the validation and testing stages (MSE = 0.56 and 0.63, respectively), being the last two values very close, which are signs of no over fitting of the models. Figure 3 shows the predicted (y-axis) and observed (x-axis) values of the 20 descriptors from the sensory analysis.

Sensory analysis of wines modeled through the bottles offers powerful tools for wineries and retailers to monitor the acceptability and quality traits of wines over time. These models can be used as quality assurance to evaluate storage conditions and the ageing process. Furthermore, these models can be used as an extra layer of information to assess wines' provenance and detect mislabeling.

https://doi.org/10.1051/bioconf/20235602003

Table 2. The regression machine learning model results in predicting wine sensory descriptors using NIR absorbance values as input. Abbreviations: R: correlation coefficient; MSE: means squared error.

Stage	Samples	Observations	R	Slope	MSE
Training	245	4900	0.97	0.93	0.26
Validatio	53	1060	0.92	0.89	0.56
Testing	53	1060	0.92	0.88	0.63
Overall	351	7020	0.95	0.91	-

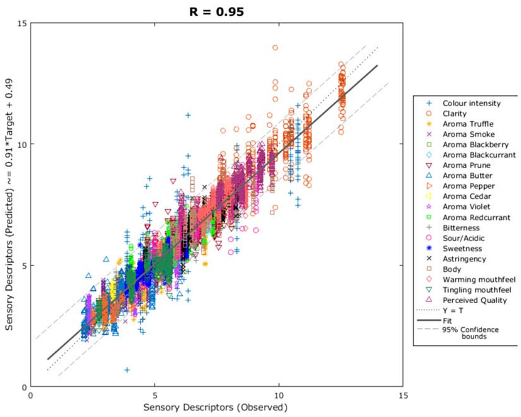

Figure 3. Overall machine learning model to predict sensory descriptors using near-infrared (NIR) absorbance values measured through the bottle as inputs.

4 Conclusion

Digital sensor technology, such as NIR spectroscopy, can effectively and accurately assess wine quality traits based on chemical fingerprinting and ML modeling with data obtained through the bottle. This novel application, in parallel with low-cost NIR devices, could be used from the winery to retailers to assess how these quality traits change during transport, aging, storage, and for the detection of tampering, adulteration, and counterfeiting. A further application can be used to assess wines' provenance and detect mislabeling.

References

1. R.K. Ranaweera, et al., Molecules **26**, 4334 (2021)
2. M.I. Rouxinol, et al., Agronomy**12**, 637 (2022)
3. H. Smyth, D. Cozzolino, Chem. Rev. **113**, 1429 (2013)
4. P. Bhardwaj, P., et al., Mach. Learn. Appl. **8**, 100261 (2022)
5. R.K. Ranaweera, et al., Food Chem. **361**, 130149 (2021)
6. M.P. Sáenz-Navajas, et al., OENO One **53**, 3 (2019)
7. N.L. da Costa, et al., Artif. Intell. Agri. **5**, 157 (2021)
8. C. Gonzalez Viejo, et al., Beverages **5**, 33 (2019)