

Physicochemical Properties Importance For Type Classification of Wines Using Machine Learning Techniques

February 1, 2022

Nuriel Shalom Mor, Ph.D.¹
Darcs and Bnei Akiva Schools, Israel

**Tigabo Asras, Eli Gal, Tesfahon Demasia, Ezra Tarab, Nathaniel Ezekiel,
Osher Nikapros, and Oshri Semimufar**
Software Engineering Specialization, Harel Bnei Akiva School, Israel

Eva Gladky, Maria Karpenko, Daniel Sason, Danis Maslov, and Omri Mor
Chemistry Specialization, Hadash Darca School, Israel

Abstract. As a subfield of artificial intelligence, machine learning designed to learn the structure of the data. Machine learning has been widely used in many scientific problems. In this study, we used machine learning techniques to figure out the most important physicochemical properties for type classification of red wines. We used a wines' dataset with 13 physicochemical properties. We used a Random Forest classifier to predict wine's type from its features, and permutation feature importance, in order to detect the most important properties of the wine for type classification. The properties: flavanoids, proline, and color intensity were found to be most important for type classification. Additional 4 classifiers: Laso classifier, Ridge classifier, Decision Tree classifier, and Support Vector classifier were used and examined for classification and feature importance. Flavanoids and proline were very important across all classifiers.

Keywords: *Portrait of wine, physicochemical properties, classifiers, permutation feature importance*

¹Please send correspondence to Nuriel S. Mor, Ph.D. Software engineering program with an emphasis on AI, Darca and Bnei Akiva schools, Israel. Email: nuriel.mor@gmail.com

1 Introduction

Today, all type of industries is improving by adopting new technologies. These technologies are also helpful to enhance the production. Wine is increasingly enjoyed by a wider range of consumers. In the last few years the consumption of wine has been increased because it has some positive correlation with heart health (Aish et al., 2018). The wine industry is investing in new technologies for the wine making and selling process (Aish et al., 2018; Cortez et al., 2009).

The purpose of this study is to find the most important physicochemical features of red wine which determine its type using machine learning techniques. Machine learning techniques excel at predicting outcomes from a set of many features (Mor & Dardeck, 2018). They are designed to find very complex relationships between input features and output, and have the ability to model complex relationships between input and output (Mor & Dardeck, 2021; Mor, 2021). Relevant to our case, machine learning techniques are suitable to learn the complex relationship between the physicochemical properties of the wine (input) to its type (output) and the importance of each property for the type classification.

We wanted to find the most important features for red wine classification from the full list of 13 physicochemical properties:

- 1) Alcohol
- 2) Malic acid
- 3) Ash
- 4) Alcalinity of ash
- 5) Magnesium
- 6) Total phenols
- 7) Flavanoids
- 8) Nonflavanoid phenols
- 9) Proanthocyanins
- 10) Color intensity
- 11) Hue
- 12) OD280/OD315 of diluted wines
- 13) Proline

To build an effective predictive model it is crucial to select the most important features that are responsible for the outcome (Chai et al., 2018).

Alcohol. An organic compound that carries at least one hydroxyl functional group ($-\text{OH}$) bound to a saturated carbon atom is alcohol. Wine can have anywhere between 5% and 23% Alcohol by Volume (ABV). The average alcohol content of wine is about 12%. This amount varies depending on the variety of wine, as well as the winemaker and their desired ABV (Vasiljevic et al., 2018)

Malic Acid. One of the main acids found in the acidity of grapes is malic acid. Its concentration decreases the more a grape ripens. Malic Acid provides a strong link to wines tasting ‘flat’ if there is not enough. If there is too much the wine will taste ‘sour’. It is vital that the levels of malic acid are monitored during the fermentation process. Quantitative determination of malic acid is important in the manufacture of wine & beer. L-Malic acid levels decrease during fermentation and in the final stages of ripening malic acid decomposes and the grape can become over ripe. Therefore, malic acid is used in the wine industry to determine the ripeness and variety of grapes. Malic acid is the principle acid found in wine and the course of malolactic fermentation is monitored by tracking the falling level of L-malic acid, and the simultaneous increasing level of L-Lactic acid (Randox Food Diagnostics, 2021).

Ash. On the average about 2.5 g/L of ash are found in wine. Ash being defined as the inorganic matter that remains after evaporation and incineration. Cations - most of the ash falls into this class and includes potassium, sodium, calcium, magnesium, iron, copper, lead, arsenic, etc. Trace minerals include pretty much anything that can be found in the soil, e.g. Aluminum, Barium, Cadmium, etc (Wine Education, 2021).

Alkalinity of ash. The alkalinity of the ash is defined as the sum of cations, other than the ammonium ion, combined with the organic acids in the wine. The alkalinity of ash will be expressed in milliequivalents per litre or in grams per litre of potassium carbonate (Institut Heidger, 2021).

Magnesium. The bioavailability of certain metal ions in grape must has been shown to be an important factor in governing fermentation performance by wine yeasts. Elevating levels of external magnesium by

supplementing growth media, or increasing intracellular concentrations of magnesium in yeast by cellular "pre-conditioning", resulted in a stimulation of yeast growth, sugar consumption rates and ethanol productivity. Elevation of calcium levels, however, tended to result in suppression of fermentation, presumably by interfering with the cellular uptake of magnesium, since the two metals are known to act antagonistically in biochemical functions. Maintenance of high magnesium: calcium concentration ratios, which are normally low in grape must, may have served to alleviate antagonism of essential magnesium-dependent yeast functions by calcium. Wine produced following fermentation with altered levels of magnesium and calcium exhibited different organoleptic profiles and implications for wine yeast physiology and wine making are discussed (Rosslyn et al., 2003).

Total Phenols. Phenol (also called carbolic acid) is an aromatic organic compound with the molecular formula C_6H_5OH . It is a white crystalline solid that is volatile. The molecule consists of a phenyl group ($-C_6H_5$) bonded to a hydroxy group ($-OH$). Mildly acidic, it requires careful handling because it can cause chemical burns. Phenolic Compounds responsible for much of the flavor and body of wine, these are a major component of wine. Benzaldehyde (vanillin) and Benzoic acid (Vanillic and Gallic acids) are the phenolic compounds one tastes the most in wines. Catechins may make up the largest quantity of Phenols. Anthocyanins are responsible for the pigmentation of red wine, and are present in proportion to the color of the wine. Resveratrol (attributed with reducing cholesterol) is a phenolic compound (Wine Education, 2021).

Flavonoids. Flavonoids are various compounds found naturally in many fruits and vegetables. They're also in plant products like wine, tea, and chocolate. There are six different types of flavonoids found in food, and each kind is broken down by your body in a different way. Flavonoids are rich in antioxidant activity and can help your body ward off everyday toxins. Including more flavonoids in your diet is a great way to help your body stay healthy and potentially decrease your risk of some chronic health conditions. Major components of red wine that have received attention as potentially cardioprotective are the flavonoids. There is mounting evidence that flavonoids and foods and beverages rich in flavonoids can make an important contribution to cardiovascular health. Fruit and vegetables, tea

and cocoa are important sources of flavonoids in the human diet. The intake of these foods has been associated with reduced risk of cardiovascular disease in population studies. Flavonoids are potent antioxidants in vitro, but it is their ability to cause vasorelaxation that is likely to be important for any vascular health benefits. Red grapes, their skin, their seeds and the wine derived from them are rich in flavonoids. Population studies have found that higher intakes of flavonoids are associated with lower risk for cardiovascular disease. In vitro studies, studies using animal models and human intervention studies have investigated how flavonoids might contribute to reduced risk of cardiovascular disease. A variety of mechanisms and outcomes have been explored and there is now strong evidence that flavonoids can improve endothelial function. A number of human studies have been performed to investigate the in vivo effects of red wine derived flavonoids on endothelial function. The results of these studies are mixed, with several studies indicating acute improvements, while other studies suggest little benefit of regular short-term consumption of red wine flavonoids. The effects of other rich dietary sources of flavonoids, such as tea and cocoa, to improve endothelial function may also provide a guide to the potential benefits of red wine flavonoids. Results of human trials and meta-analyses of these trials suggest that tea, cocoa and flavonoid-rich fruits can improve endothelial function. Therefore there is good evidence that flavonoid-rich foods and beverages can have vascular health benefits. However, because bioactivity of different flavonoids varies, health effects cannot be generalized to all flavonoids and flavonoid-rich foods. Further studies are needed to establish any vascular health benefits of the red wine flavonoids (Hodgson, 2014).

Nonflavanoid phenols. As their name suggests, non-flavonoids include most of the small phenolic compounds in wine that are not flavonoids, such as hydroxycinnamates, hydroxybenzoic acids, and stilbenes. The hydroxycinnamates are found in all plants and include caftaric, coutaric, and fertaric acid. They're found in grape pulp and are the most abundant non-flavonoids in wine. Hydroxycinnamates act as cofactors in copigmentation and participate in oxidation and juice browning. These are also the most important phenolic compounds in white wines (A Guide to Wine Phenolics, 2021)

Proanthocyanidins. They give the fruit or flowers of many plants their red, blue, or purple colors. They are a class of polyphenols found in many

plants, such as cranberry, blueberry, and grape seeds. Chemically, they are oligomeric flavonoids. Many are oligomers of catechin and epicatechin and their gallic acid esters. More complex polyphenols, having the same polymeric building block, form the group of tannins. Polyphenols are a large family of naturally occurring organic compounds characterized by multiples of phenol units. They are abundant in plants and structurally diverse. Polyphenols include flavonoids, tannic acid, and ellagitannin, some of which have been used historically as dyes and for tanning garments. Proanthocyanidins play an important role in wine; with the capability to bind salivary proteins, these condensed tannins strongly influence the perceived astringency of the wine. These compounds are typically present in levels of 300mg/L in red wine, though enological processing can affect the final concentrations. Proanthocyanidins are constructed from monomeric flavan-3-ols, which is considered one of the largest and most functional subclass of flavonoids found in foods and beverages² (Health Encyclopedia, 2021).

Color intensity. The intensity of color can be observed with the wine's opacity. Deeply opaque red wines have been noted for having more pigment and phenolics than more translucent red wines. For example, Syrah has as much as 4 times more pigment (antioxidants) than Zinfandel. There are a few features you can observe that are generally true with color intensity: Different grape varieties have different levels of intensity. For example, Gamay is very low and Pinotage has exceptionally high levels of pigmentation. Color intensity can be amplified by other polyphenols (e.g. tannin) in wine. Thus, wines that are more opaque may also contain higher levels of tannin. The pigment in red wine is sensitive to both temperature and sulfites. Wines that are fermented at high temperatures or have higher sulfur additions will have less color intensity. Wines lose pigment as they age. As much as 85% of the anthocyanin is lost after 5 years (Secrets Behind the Color Pigment in Red Wine, 2021).

Hue. If you look at a red wine under natural lighting conditions and over a white background, you'll get a pretty accurate impression of its hue. It might be difficult to see at first, but young red wines (under 5 years) range in hue from red, to violet, to blue. You can see this hue by looking towards the edge of the wine as it hits the glass. Wines with more red colored hue have a lower pH (high acidity). Wines with a violet colored hue range from

around 3.4–3.6 pH (on average). Wines with a more blueish tint (almost like magenta) are over 3.6 pH and possibly closer to 4 (low acidity). Of course, each red grape variety expresses color a little bit differently and there are many variables that will affect the color (variables such as co-pigmentation, sulfur additions, etc.), but the above is generally true. Examples: Malbec: A highly tinted red wine that, when produced in a soft and lush style, often has a magenta (blue) tint on the rim of the glass. Sangiovese: A less tinted red wine (often translucent) that’s spicy character is partially explained by high acidity, which you can see in its brilliant red color. (Secrets Behind the Color Pigment in Red Wine, 2021).

OD280/OD315 of diluted wines. This is a method for determining the protein concentration, which can determine the protein content of various wines (Bai et al., 2019).

Proline. The most abundant amino acid present in grape juice and wine is proline. The amount present is influenced by viticultural and winemaking factors and can be of diagnostic importance. A method for rapid routine quantitation of proline would therefore be of benefit for wine researchers and the industry in general (Long et al., 2012).

2 Method

The original dataset contained the following features:

- 1) Alcohol
- 2) Malic acid
- 3) Ash
- 4) Alcalinity of ash
- 5) Magnesium
- 6) Total phenols
- 7) Flavanoids
- 8) Nonflavanoid phenols
- 9) Proanthocyanins
- 10) Color intensity
- 11) Hue
- 12) OD280/OD315 of diluted wines
- 13) Proline

The data is the results of a chemical analysis of wines grown in the same region in Italy by three different cultivators. There are thirteen different measurements taken for different constituents found in the three types of wine.

The data used in this experiment was obtained from the UCI database of the Italian wine data set, which contained a sample size of 178. The data contained in each variable is the result of chemical analysis. The Italian wines shown in the sample are grown in the same area but from different varieties.

<https://archive.ics.uci.edu/ml/datasets/wine>

As mentioned, the data set consists of a total of 13 numeric variables:

(1) Malic acid: It is a kind of acid with strong acidity and apple aroma. The red wine is naturally accompanied by malic acid. (2) Ash: The essence of ash is an inorganic salt, which has an effect on the overall flavor of the wine and can give the wine a fresh feeling. (3) Alkalinity of ash: It is a measure of weak alkalinity dissolved in water. (4) Magnesium: It is an essential element of the human body, which can promote energy metabolism and is weakly alkaline. (5) Total phenols: molecules containing polyphenolic substances, which have a bitter taste and affect the taste, color and taste of the wine, and belong to the nutrients in the wine. (6) Flavanoids: It is a beneficial antioxidant for the heart and anti-aging, rich in aroma and bitter. (7) Nonflavanoid phenols: It is a special aromatic gas with oxidation resistance and is weakly acidic. (8) Proanthocyanins: It is a bioflavonoid compound, which is also a natural antioxidant with a slight bitter smell. (9) Color intensity: refers to the degree of color shade. It is used to measure the style of wine to be "light" or "thick". The color intensity is high, meanwhile the longer the wine and grape juice are in contact during the wine making process, the thicker the taste. (10) Hue: refers to the vividness of the color and the degree of warmth and coldness. It can be used to measure the variety and age of the wine. Red wines with higher ages will have a yellow hue and increased transparency. Color intensity and hue are important indicators for evaluating the quality of a wine's appearance. (11) Proline: It is the main amino acid in red wine and an important part of the nutrition and flavor of wine. (12) OD280/OD315 of diluted wines: This is a method for determining the protein concentration, which can determine the protein content of various wines (Bai et al., 2019).

We used a Random Forest classifier to predict wine’s type from these 13 features. A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting (Dogru et., 2018). Decision Trees are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A tree can be seen as a piecewise constant approximation (Myles et al., 2014).

To perform the model inspection technique known as permutation feature importance we used scikit-learn’s built-in function called `permutation_importance`. Permutation feature importance is a model inspection technique that can be used for any fitted estimator when the data is tabular. This is especially useful for non-linear or opaque estimators. The permutation feature importance is defined to be the decrease in a model score when a single feature value is randomly shuffled. This procedure breaks the relationship between the feature and the target, thus the drop in the model score is indicative of how much the model depends on the feature. This technique benefits from being model agnostic and can be calculated many times with different permutations of the feature (Breiman, 2001).

The importance score is computed in a way that higher values represent better predictive power.

3 Experiments and Results

Mean Accuracy score obtained using Random Forest classifier:
91.11%

The importance scores (`permutation_importance`):

Feature **flavanoids** with index 6 has an average importance score of 0.227
+/- 0.025
Feature **proline** with index 12 has an average importance score of 0.142
+/- 0.019

Feature **color_intensity** with index 9 has an average importance score of 0.112 +/- 0.023

Feature **od280/od315_of_diluted_wines** with index 11 has an average importance score of 0.007 +/- 0.005

Feature **total_phenols** with index 5 has an average importance score of 0.003 +/- 0.004

Feature **malic_acid** with index 1 has an average importance score of 0.002 +/- 0.004

Feature **proanthocyanins** with index 8 has an average importance score of 0.002 +/- 0.003

Feature **hue** with index 10 has an average importance score of 0.002 +/- 0.003

Feature **nonflavanoid_phenols** with index 7 has an average importance score of 0.000 +/- 0.000

Feature **magnesium** with index 4 has an average importance score of 0.000 +/- 0.000

Feature **alkalinity_of_ash** with index 3 has an average importance score of 0.000 +/- 0.000

Feature **ash** with index 2 has an average importance score of 0.000 +/- 0.000

Feature **alcohol** with index 0 has an average importance score of 0.000 +/- 0.000

In the test test we found:

Give an indication for generalization.

Feature **flavanoids** with index 6 has an average importance score of 0.202 +/- 0.047

Feature **proline** with index 12 has an average importance score of 0.143 +/- 0.042

Feature **color_intensity** with index 9 has an average importance score of 0.112 +/- 0.043

Feature **alcohol** with index 0 has an average importance score of 0.024 +/- 0.017

Feature **magnesium** with index 4 has an average importance score of 0.021 +/- 0.015

Feature **od280/od315_of_diluted_wines** with index 11 has an average importance score of 0.015 ± 0.018

Feature **hue** with index 10 has an average importance score of 0.013 ± 0.018

Feature **total_phenols** with index 5 has an average importance score of 0.002 ± 0.016

Feature **nonflavanoid_phenols** with index 7 has an average importance score of 0.000 ± 0.000

Feature **alkalinity_of_ash** with index 3 has an average importance score of 0.000 ± 0.000

Feature **malic_acid** with index 1 has an average importance score of -0.002 ± 0.017

Feature **ash** with index 2 has an average importance score of -0.003 ± 0.008

Feature **proanthocyanins** with index 8 has an average importance score of -0.021 ± 0.020

If a feature is deemed as important for the train set but not for the testing, this feature will probably cause the model to overfit.

3.1 Retraining With The Most Important Properties

We re-trained the Random Forest classifier with only the top 3 most important features.

On TRAIN split:

Feature **flavanoids** with index 6 has an average importance score of 0.227 ± 0.025

Feature **proline** with index 12 has an average importance score of 0.142 ± 0.019

Feature **color_intensity** with index 9 has an average importance score of 0.112 ± 0.023

On TEST split:

Feature **flavanoids** with index 6 has an average importance score of 0.202 ± 0.047

Feature **proline** with index 12 has an average importance score of 0.143 +/- 0.042

Feature **color_intensity** with index 9 has an average importance score of 0.112 +/- 0.043

Mean Accuracy score obtained using Random Forest classifier and 3 features: 93.33%

By using only the 3 most important features the model achieved a mean accuracy even higher than the one using all 13 features!

3.2 Results Obtained From Different Classifiers

Lasso classifier

Least absolute shrinkage and selection operator (LASSO) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model. Lasso was originally formulated for linear regression models (Galla, 2020).

Mean accuracy score on the test set: 86.80%

Top 4 features when using the test set:

Feature **flavanoids** with index 6 has an average importance score of 0.323 +/- 0.055

Feature **proline** with index 12 has an average importance score of 0.203 +/- 0.035

Feature **od280/od315_of_diluted_wines** with index 11 has an average importance score of 0.146 +/- 0.030

Feature **alcalinity_of_ash** with index 3 has an average importance score of 0.038 +/- 0.014

Ridge classifier

In ridge regression, the cost function is altered by adding a penalty equivalent to the square of the magnitude of the coefficients. Ridge regression shrinks the coefficients and it helps to reduce the model complexity and multi-collinearity (Fan et al., 2021).

Mean accuracy score on the test set: 88.71%

Top 4 features when using the test set:

Feature **flavanoids** with index 6 has an average importance score of 0.445
+/- 0.071

Feature **proline** with index 12 has an average importance score of 0.210
+/- 0.035

Feature **color_intensity** with index 9 has an average importance score of
0.119 +/- 0.029

Feature **od280/od315_of_diluted_wines** with index 11 has an average
importance score of 0.111 +/- 0.026

Decision Tree classifier

Decision Trees are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features (Myles et al., 2014).

Mean accuracy score on the test set: 95.56%

Top 4 features when using the test set:

Feature **flavanoids** with index 6 has an average importance score of 0.297
+/- 0.061

Feature **proline** with index 12 has an average importance score of 0.143
+/- 0.039

Feature **color_intensity** with index 9 has an average importance score of
0.131 +/- 0.037

Feature **alcohol** with index 0 has an average importance score of 0.049 +/-
0.020

Support Vector classifier

The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence (Galla et al., 2020).

Mean accuracy score on the test set: 97.78%

Top 4 features when using the test set:

Feature **proline** with index 12 has an average importance score of 0.069
+/- 0.031

Feature **flavanoids** with index 6 has an average importance score of 0.061
+/- 0.023

Feature **alcohol** with index 0 has an average importance score of 0.044 +/-
0.023

Feature **ash** with index 2 has an average importance score of 0.032 +/-
0.018

Looks like flavanoids and proline are very important across all classifiers.
However there is variability from one classifier to the others on what
features are considered the most important ones.

4 Conclusion

By using only the 3 most important features the model achieved a mean
accuracy even higher than the one using all 13 features!

Looks like flavanoids and proline are very important across all classifiers.
However there is variability from one classifier to the others on what
features are considered the most important ones.

Notebook

5 References

A Guide to Wine Phenolics (2021). [Link](#)

Aich, S., Al-Absi, A. A., Hui, K. L., Lee, J. T., & Sain, M. (2018, February). A classification approach with different feature sets to predict the quality of different types of wine using machine learning techniques. In 2018 20th International conference on advanced communication technology (ICACT) (pp. 139-143). IEEE.

- Bai, X., Wang, L., & Li, H. (2019). Identification of red wine categories based on physicochemical properties. In International Conference on Education Technology, Management and Humanities Science (ETMHS 2019) (pp. 1479-1484).
- Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, 300, 70-79
- Cortez, P., Teixeira, J., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009, October). Using data mining for wine quality assessment. In International Conference on Discovery Science (pp. 66-79). Springer, Berlin, Heidelberg.
- Dogru, N., & Subasi, A. (2018, February). Traffic accident detection using random forest classifier. In 2018 15th learning and technology conference (L&T) (pp. 40-45). IEEE.
- Fan, P., Deng, R., Qiu, J., Zhao, Z., & Wu, S. (2021). Well logging curve reconstruction based on kernel ridge regression. *Arabian Journal of Geosciences*, 14(16), 1-10.
- Forina, M. et al, PARVUS - An Extendible Package for Data Exploration, Classification and Correlation. Institute of Pharmaceutical and Food Analysis and Technologies, Via Brigata Salerno, 16147 Genoa, Italy.
- Galla, D. K. K., Mukamalla, B. R., & Chegiredy, R. P. R. (2020). Support vector machine based feature extraction for gender recognition from objects using lasso classifier. *Journal of Big Data*, 7(1), 1-16.
- Health Encyclopedia (2021). [Link](#)
- Hodgson, J. M. (2014). Red wine flavonoids and vascular health. *Nutrition and Aging*, 2(2, 3), 139-144.
- L. Breiman, "Random Forests", *Machine Learning*, 45(1), 5-32, 2001.
- Long, D., Wilkinson, K. L., Poole, K., Taylor, D. K., Warren, T., Astorga, A. M., & Jiranek, V. (2012). Rapid method for proline determination in grape juice and wine. *Journal of agricultural and food chemistry*, 60(17), 4259-4264.

Institut Heidger (2021).
Link

Lichman, M. (2013). UCI Machine Learning. Repository
[<https://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School
of Information and Computer Science.

Mor, N. S., & Dardeck, K. L. (2021). Applying a Convolutional Neural
Network to Screen for Specific Learning Disorder. *Learning Disabilities: A
Contemporary Journal*, 19(2), 161-169.

Mor, N. S., & Dardeck, K. L. (2020). Applying Deep Learning to Specific
Learning Disorder Screening. *arXiv preprint arXiv:2008.13525*.

Mor, N. S., & Dardeck, K. L. (2018). Quantitative Forecasting of Risk for
PTSD Using Ecological Factors: A Deep Learning Application. *Journal of
Social, Behavioral, and Health Sciences*, 12(1), 61-73.
DOI:10.5590/JSBHS.2018.12.1.04.

Mor, N. S. (2021). Generating Photo-realistic Images from LiDAR Point
Clouds with Generative Adversarial Networks. *arXiv preprint
arXiv:2112.11245*.

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004).
An introduction to decision tree modeling. *Journal of Chemometrics: A
Journal of the Chemometrics Society*, 18(6), 275-285.

Randox Food Diagnostics (2021). Link

Rosslyn M. Birch, Maurizio Ciani & Graeme M. Walker (2003) Magnesium,
Calcium and Fermentative Metabolism in Wine Yeasts, *Journal of Wine
Research*, 14:1, 3-15, DOI: 10.1080/0957126032000114973

S. Aeberhard, D. Coomans and O. de Vel, Comparison of Classifiers in High
Dimensional Settings, Tech. Rep. no. 92-02, (1992), Dept. of Computer
Science and Dept. of Mathematics and Statistics, James Cook University of
North Queensland. (Also submitted to Technometrics).

The data was used with many others for comparing various classifiers. The classes are separable, though only RDA has achieved 100% correct classification. (RDA : 100%, QDA 99.4%, LDA 98.9%, 1NN 96.1% (z-transformed data)) (All results using the leave-one-out technique)

S. Aeberhard, D. Coomans and O. de Vel, "THE CLASSIFICATION PERFORMANCE OF RDA" Tech. Rep. no. 92-01, (1992), Dept. of Computer Science and Dept. of Mathematics and Statistics, James Cook University of North Queensland. (Also submitted to Journal of Chemometrics).

Secrets Behind the Color Pigment in Red Wine (2021).
[https://winefolly.com/deep-dive/
know-a-wine-just-by-looking-at-the-color/](https://winefolly.com/deep-dive/know-a-wine-just-by-looking-at-the-color/)

Vasiljevic, M., Couturier, D. L., & Marteau, T. M. (2018). Impact on product appeal of labeling wine and beer with (a) lower strength alcohol verbal descriptors and (b) percent alcohol by volume (% ABV): An experimental study. *Psychology of Addictive Behaviors*, 32(7), 779.

Wine Education (2021).
<http://www.wineeducation.com/wineismadeof.html>