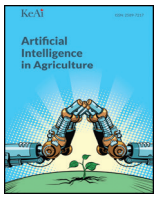
Artificial Intelligence in Agriculture 7 (2023) 58–68

Contents lists available at ScienceDirect

Artificial Intelligence in Agriculture

journal homepage: http://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/



t-SNE: A study on reducing the dimensionality of hyperspectral data for the regression problem of estimating oenological parameters

Rui Silva [a,*], Pedro Melo-Pinto [a,b,**]

[a] CITAB - Centre for the Research and Technology of Agro-Environmental and Biological Sciences, Inov4Agro-Institute for Innovation, Capacity Building and Sustainability of Agri-Food Production, Universidade de Trás-os-Montes e Alto Douro, Quinta dos Prados, Vila Real 5000-801, Portugal
[b] Departamento de Engenharias, Escola de Ciências e Tecnologia, Universidade de Trás-os-Montes e Alto Douro, Quinta dos Prados, Vila Real 5000-801, Portugal

ARTICLE INFO

Article history:
Received 22 September 2022
Received in revised form 20 February 2023
Accepted 21 February 2023
Available online 6 March 2023

Keywords:
Hyperspectral images
Dimensionality reduction
Regression
T-SNE
Support vector machines
Wine grape berries

ABSTRACT

In recent years there is a growing importance in using machine learning techniques to improve procedures in precision agriculture: in this work we perform a study on models capable of predicting oenological parameters from hyperspectral images of wine grape berries, a specially relevant topic to boost production tasks for winemakers. Specifically, we explore the capabilities of a novel technique mostly used for visualization, t-Distributed Stochastic Neighbor Embedding (t-SNE), for reducing the dimensionality of the highly complex hyperspectral data and compare its performance with Principal Component Analysis (PCA) method, which despite the introduction of many nonlinear dimensionality reduction techniques over the years, had achieved the best results for real-world data across several studies in literature. Additionally we explore the potential of Kernel t-SNE, an extension to the t-SNE method that allows for the usage of the technique in streaming data or online scenarios. Our results show that, in a direct comparison, t-SNE achieves better metrics than PCA for most of the data sets in this work and that the regressor (Support Vector Regression, SVR) performs better with the t-SNE reduced features as inputs, accomplishing better predictions with lower error rates. Comparing the results with current literature, our shallow learning model paired with t-SNE achieves either better or on par results than those reported, even competing with more advanced models that use deep learning techniques, which should propel the introduction of t-SNE in more studies that require dimensionality reduction.

1. Introduction

In recent years the wine industry has evolved in a way to introduce as many new technologies as possible to improve production procedures: one of its main focuses is to obtain grapes for wine production in an environmentally friendly manner, without destroying them in the process and choosing them according to quality features. The traditional approach relies on laboratorial analysis to assess some select oenological parameters which, alongside destroying the grapes used for analysis, is a cost and time-heavy method. The logical next step was to find some sort of imaging technique that allowed to obtain clean information about the grapes, and hyperspectral imaging found the most success in this task.

Hyperspectral imaging (Gowen et al., 2007; Hall et al., 2002) is a technique that shows the light reflection and absorption on an object as a function of wavelength: it collects both spatial and spectral information, requiring robust models that are capable of extracting knowledge from the patterns present in the spectra. Our recent studies (Fernandes et al., 2011, 2015; Gomes et al., 2014a, 2014b, 2017b, 2021a, 2021b; Gomes and Melo-Pinto, 2021; Silva et al., 2018) found some success in combining these images with machine/deep learning algorithms that are capable of regression from hyperspectral images; however, one of the biggest issues to tackle in order to obtain generalization capacity from these models is the ability to reduce the input space, which allows to consistently identify the main features in the spectra and accurately predict oenological parameters. This led to a study of a wide variety of dimensionality reduction methods (Silva and Melo-Pinto, 2021) that showed that despite several advances and new techniques, Principal Component Analysis (PCA) (Wold et al., 1987) was still the method that allowed the algorithm to achieve the best predictions and the best generalization capacity for the case of predicting oenological parameters from hyperspectral images of wine grape berries. Following this line of research, arose the need of applying

* Corresponding author.
** Corresponding author at: CITAB - Centre for the Research and Technology of Agro-Environmental and Biological Sciences, Inov4Agro-Institute for Innovation, Capacity Building and Sustainability of Agri-Food Production, Universidade de Trás-os-Montes e Alto Douro, Quinta dos Prados, Vila Real 5000-801, Portugal.
E-mail addresses: ruimsilva@utad.pt (R. Silva), pmelo@utad.pt (P. Melo-Pinto).

https://doi.org/10.1016/j.aiia.2023.02.003

the t-Distributed Stochastic Neighbor Embedding (t-SNE) technique for the dimensionality reduction of hyperspectral images.

t-SNE (Van der Maaten and Hinton, 2008) is a technique that visualises high-dimensional data by giving each data point a location in a two or three-dimensional map, reducing the tendency to crowd points together and therefore creating more structured visualisations of the data. We decided to conduct a study of the application of t-SNE to hyperspectral imaging data since, as concluded in (Silva and Melo-Pinto, 2021) and mentioned in (Van der Maaten and Hinton, 2008), other dimensionality reduction techniques showcase a strong performance on artificial data sets but fail to translate that performance to real-world data, mostly because of their failure to retain both the local and global structure of the data in a single map: t-SNE has shown to be capable of capturing much of the local structure of the high dimensional data as well as revealing a global structure in it.

Furthermore, it is also mentioned in (Van der Maaten and Hinton, 2008) that it is unclear how t-SNE will perform on general dimensionality reduction tasks, and we decided to evaluate its performance on hyperspectral imaging data - reviewing the current state-of-the-art shows some applications of t-SNE to hyperspectral data:

- in (Miao et al., 2018) the authors use t-SNE to reduce the dimensionality of hyperspectral images of maize kernels and perform classification.
- in (Hariharan, 2021) t-SNE is applied to reduce the dimensionality of hyperspectral open-access data sets (i.e.: Indian Pines data set) and perform classification while tackling the curse of high dimensionality on a limited number of training samples.
- in (Zhang et al., 2018) t-SNE is combined with a Deep Convolutional Generative Adversarial Network (DCGAN) to extract spectra-spatial features and perform dimensionality reduction on hyperspectral images.
- in (Devassy and George, 2020) t-SNE is used to perform clustering and obtain a better visualization on a hyperspectral database of inks from 60 different pens.
- in (Gao et al., 2019) the authors combine t-SNE with a Convolutional Neural Network (CNN) to reduce the dimensionality and perform classification on hyperspectral images with a large number of bands but an insufficient number of sample pixels for each class.
- in (Pouyet et al., 2018) t-SNE is also used to obtain visualisations of hyperspectral images in 2D scatter plots.

Additionally, there are also relevant applications of t-SNE outside of hyperspectral images:

- in (Gisbrecht et al., 2012) the authors pave the way towards efficient nonlinear dimensionality reduction proposing an extension of t-SNE with the linear basis function.
- in (Alibert, 2019) t-SNE is applied to better visualise and represent planetary systems in a two-dimensional space.
- in (Anowar et al., 2021) the authors compare the performance of several dimensionality reduction methods on open-access data sets.

However, and as far as we are aware up to date, there is not a study of the capability of t-SNE to reduce the dimensionality of a data set to then perform predictions with a machine learning algorithm (regression problem); additionally, we also perform this study on real-world samples (most of the studies are performed on artificial data sets) of hyperspectral images of wine grape berries, a problem with extremely high variability between harvest years and vintages of wine grapes; our study is also relevant because we perform t-SNE on small data sets (due to the inherent difficulties in acquiring training samples) in contrast to most works that have great amounts of data: combining all these aspects, we believe we perform a true test to the capability of t-SNE to reduce the dimensionality of real-world hyperspectral images in a problem with real-world applications. To further enhance our study of the t-SNE technique, we also decided to apply a variation of this method, Kernel t-SNE. Kernel t-SNE (Gisbrecht et al., 2015) extends the non-parametric dimensionality reduction technique to an explicit mapping by fixing the parametric form $x \rightarrow f_w(x) = y$ and optimizing the parameters of f_w instead of the projection coordinates: this enables to map large data sets in linear time by training a mapping on a small sub-sample only, with good generalization capacity. However, since Kernel t-SNE is applied to a subset of the training samples only, the results may differ when compared to t-SNE that is applied to the full data set, due to missing information in the data used for training of the map. In order to close this information gap, the authors (Schulz and Hammer, 2015) introduced Fisher Kernel t-SNE: a set of data points x_i is equipped with the pairwise Fisher metric, estimated based on the class labels taking simple linear approximations for the path integrals and, using t-SNE, a new training set X' is obtained by taking the auxiliary label information into account (the calculated pairwise distances of data computed based on the Fisher metric); the authors then infer a kernel t-SNE mapping which is adapted to the label information due to the information inherent in the training set, and the resulting map is adapted to the information encoded in the training set - this technique can be referred to as Fisher t-SNE or Fisher kernel t-SNE. Both these techniques attempt to further optimize the capability of t-SNE in real-world applications and while we decided to study kernel t-SNE, we opted to not implement Fisher kernel-SNE due to the high computational cost of calculating the Fisher information matrices, which renders the solution unfitting for a possible real-time analysis in the vineyards.

As for the rest of this paper, in Sub-Section 2.1 we describe the hyperspectral imaging procedure, with an in-depth look into our experimental setup and to the way we perform the reflectance measurements to construct the training sets; Sub-Section 2.2 provides a brief theoretical background of the dimensionality reduction methods applied; Sub-Section 2.3 introduces the algorithm chosen to perform regression, the Support Vector Regression (SVR) technique; in Sub-Section 2.4 we discuss our approach to avoid over-fitting and achieve maximum generalization capacity via the usage of a cross-validation technique; in Sub-Section 2.5 we give insights about the grape sampling procedure and provide the data sets description to give a better understanding of the data and its high variability; Section 3 presents the results obtained for the prediction of the oenological parameters for each dimensionality reduction technique and for each oenological parameter, alongside a discussion of said results and how they fare when compared to other state-of-the-art publications; Section 4 summarizes our findings and concludes about the work, while also denoting future guidelines for improvement.

2. Materials and methods

2.1. Hyperspectral images

2.1.1. Experimental setup

Hyperspectral measurements were performed in line with our previous work (Silva and Melo-Pinto, 2021) using the following image acquisition system (Fig. 1): a hyperspectral camera, composed of a JAI Pulnix (JAI, Yokohama, Japan) black and white camera and a Specim Imspector V10E spectrograph (Specim, Oulu, Finland); lighting, by means of a lamp holder with $300 \times 300 \times 175\ mm^3$ (length × width × height) that held four 20 W, 12 V halogen lamps and two 40W, 220V blue reflector lamps (Spotline, Philips, Eindhoven, The Netherlands). The halogen lamps were powered by continuous current power supplies to avoid light flickering and the reflector lamps were powered at only 110V to reduce lighting and prevent camera saturation. Each component was bough directly from their respective manufacturers and the image acquisition system was assemble by the authors. The resulting images have a spatial resolution of 1040 × 1392 pixels, where the 1040 pixels correspond to the wavelength channels, ranging between 380 and 1028 nm, with approximately 0.6 nm width for each channel,

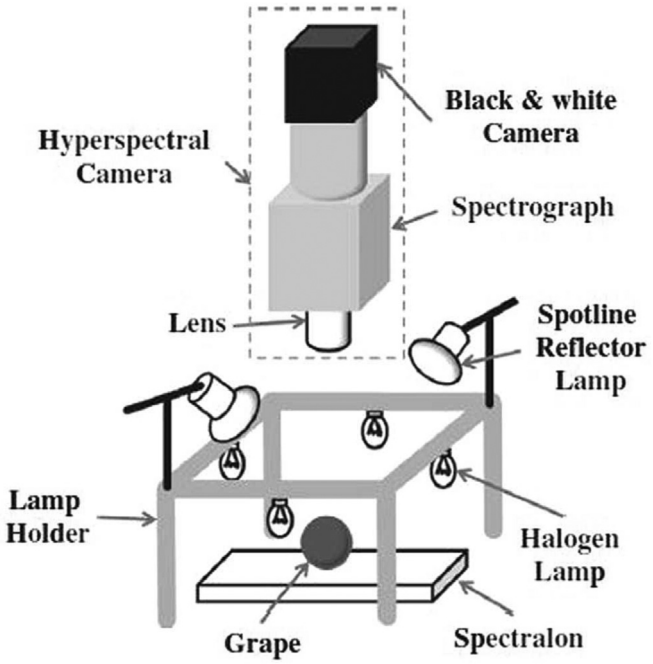


Fig. 1. Mock-up of the setup used for hyperspectral imaging (Fernandes et al., 2011).

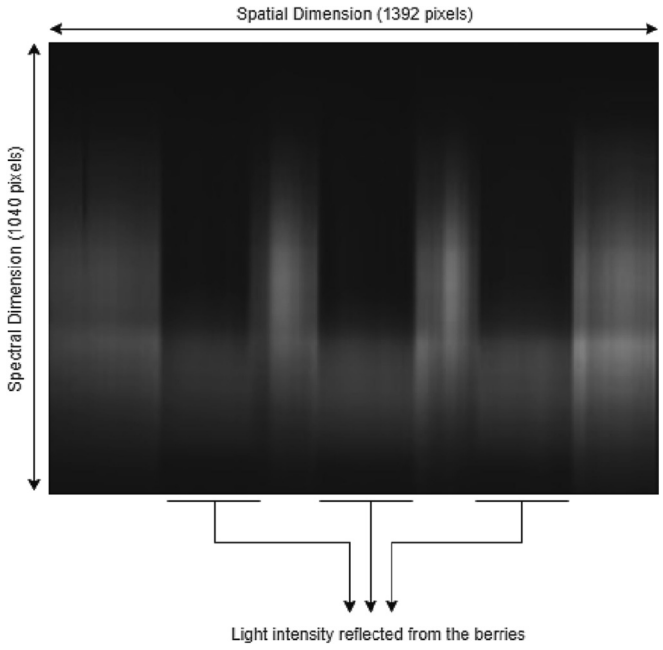


Fig. 3. Hyperspectral image of a wine grape berry sample before segmentation and reflectance measurements.

and the 1392 pixels stand for the spatial dimension (one line over the samples) with approximately 110 *mm* of width. The distance between the camera and the sample base was set to 420 *mm* and the camera was controlled with the Coyote software from JAI. All the hyperspectral measurements were done inside a dark room and at room temperature (20 °C).

Each hyperspectral image is acquired for six grape berries and in three different berry rotations of approximately 120° between positions, with a single line taken solely over the berry equator when considering the pedicel as the pole, as seen below in Fig. 2.

The final hyperspectral image for each rotation and position is the average of 32 different hyperspectral images acquired during a period of time of 4 s, with the camera acquiring 8 images per second: this method reduces measurement noise, and also provides with some information of the spatial dimensions (since we are not using it directly on line-scan hyperspectral images) that is reflected on the averaged value. As expected, there are observable patterns of light reflection and absorption in the main areas where the wine grape berries lie in the Spectralon (full reflection surface) and, to get individual images for each wine grape berry, we use a threshold-based segmentation method. Fig. 3 shows an example of a hyperspectral image captured by the aforementioned experimental setup before segmentation.

Observing Fig. 3 it is possible to denote that the spots where the grapes lie for imaging are clear to absorb more light (black strips) while the spots with no grapes (the remainder of the Spectralon) reflect more light (white strips). After finishing the process of obtaining an individual hyperspectral image for each wine grape berry we carry out reflectance measurements to build the final data sets.

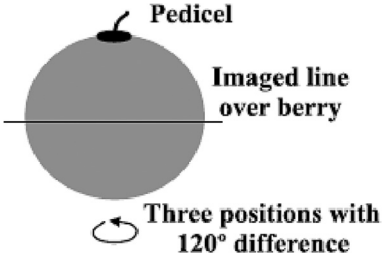


Fig. 2. Imaging line over each berry (Gomes et al., 2017a).

2.1.2. Reflectance measurements

Reflectance is a function of the light wavelength and is defined as the ratio between the light intensity reflected by an object and the light that illuminates that object. Despite the fact that the measurements could be carried out using other modalities, like transmittance or interactance, we chose the reflectance mode as input because the different patterns of reflectance and absorption across wavelengths will allow for the identification of chemical compounds and because, contrary to the other referred modes, imaging is possible without the need for the spectrometer/camera to touch the samples (Gomes et al., 2021a, 2021b; Gomes and Melo-Pinto, 2021; Silva and Melo-Pinto, 2021). For a position represented by vector x and at wavelength λ, the reflectance R can be expressed as:

$$R(x,\lambda) = \frac{\alpha(x,\lambda) - \sigma(x,\lambda)}{\mu(x,\lambda) - \sigma(x,\lambda)} \quad (1)$$

where α is the light intensity reflected from the grape; μ is the light intensity reflected from a reference total reflectance target; and σ is the dark current signal, which is electronic noise. For each wine grape berry we took 32 hyperspectral images, with three different berry rotations, and the resulting spectrum was then normalized (using max-min normalization) to avoid variations in the measured light intensities. Fig. 4 shows the result of the reflectance measurements in one of the data sets that will be used in the present work.

2.2. Dimensionality reduction

2.2.1. t-distributed stochastic neighbor embedding

t-SNE (Van der Maaten and Hinton, 2008) is a technique that is capable of capturing the local structure of the high-dimensional data while also revealing global structure, such as the presence of clusters at several scales. When a problem requires dimensionality reduction common goals between different applications can be highlighted, such as

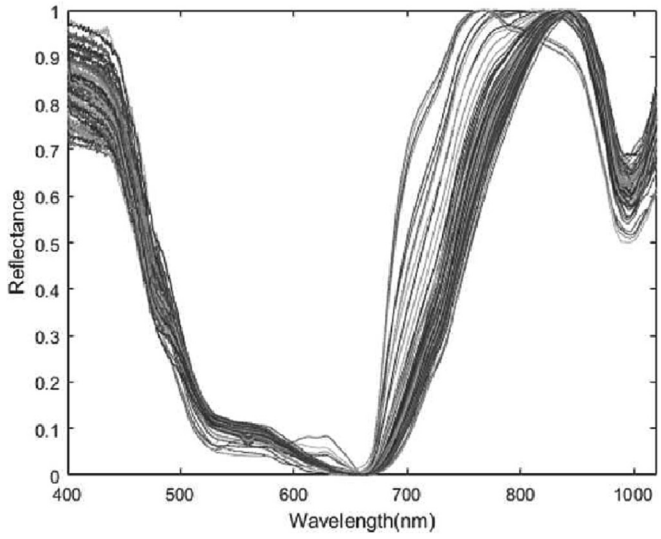


Fig. 4. Reflectance values obtained for a complete data set (Silva and Melo-Pinto, 2021).

preserving as much significant structure or information in the high-dimensional data as in the low-dimensional representation/projection; increase the means of interpretation of the data in the lower dimension; and minimize the information loss in the new, low-dimensional representation of the data. In our previous studies (Silva and Melo-Pinto, 2021) PCA established itself as the technique with the best results in terms of easing the interpretation of the data by the regressor, obtaining the representations that contained the most preserved information and, in most cases, the representations that led to less erroneous predictions by the machine learning algorithm. When we compare these two techniques, we can denote some relevant differences:

- PCA is a deterministic algorithm commonly used for feature extraction, while t-SNE is a randomized algorithm that is mostly used for visualization purposes only;
- PCA applies a linear technique where the focus is on keeping the dissimilar points apart in a lower-dimensional space, while t-SNE applies a non-linear technique that attempts to keep the similar data points close together in the lower-dimensional space;
- PCA transforms the original data by preserving the variance in the data using eigenvalues matrices and is very affected by outliers, while t-SNE preserves the local structure of the data by using student t-distributions to compute the similarity between two points in the lower-dimensional space (which helps address crowding and optimizations problems) and is not as strongly impacted by outliers.

Henceforth, since t-SNE is mostly used for visualization purposes it is unclear how it will perform in a general dimensionality reduction task, and by having such contrasting characteristics to the technique that obtained the best results thus far, we were intrigued to conduct this application and study the outcome since we concluded previously (Silva and Melo-Pinto, 2021) that, in non-linear techniques, local learners seemed to have better results than global learners, but we never tested a non-linear technique capable of preserving both local and global structure; and also, since the problem of estimating oenological parameters from wine grape berries possesses high variability due to the hundreds of different varieties and the different harvest years, a technique capable of reducing the impact of outliers could lead to a superior generalization capacity.

t-SNE originates from Stochastic Neighbor Embedding (SNE) (Hinton and Roweis, 2002), but uses a student t-distribution with a heavy-tailed probability distribution to address the crowding problem in the original technique: in summary, t-SNE minimizes the Kullback-Leibler divergences between the high-dimensional and the latent spaces with the cost function:

$$C = \sum_i KL(P_i \| Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}} \tag{2}$$

where P is the conditional probability distribution in the original space and Q_i is the conditional probability distribution in the latent space. Since the Kullback-Leibler divergence is not symmetric, the error in the pairwise distances in the low-dimensional representation will be weighted differently: the usage of widely separated map points to represent nearby data points will have a larger cost while the usage of nearby map points to represent widely separated data points will have a smaller cost - this means that there is a focus on retaining the local structure of the data in the low-dimensional representation.

Hence, the variance of the Gaussian that is centered on data point x_i, the parameter σ_i, will never be optimally represented for all data points in the data set because the density of the data is likely to vary: in denser regions a smaller value of σ will be more appropriate than in sparser regions. t-SNE will then perform a binary search for the value of σ_i that produces a P_i with a fixed perplexity that is user-specified and defined as:

$$\text{Perp}\,(P_i) = 2^{H(P_i)} \tag{3}$$

where $H(P_i)$ is the Shannon entropy of P_i measured in bits. One can then interpret perplexity as a simple measure of the effective number of neighbors with typical values varying between 5 and 50: in this work, we used the same range of values for all the experiments, picking the perplexity value that originated the predictions with the least error value.

2.2.2. Kernel t-SNE

t-SNE is a non-parametric technique that provides a high-to-low dimensional representation of a given data set without a mapping formula on how to project further points not included in the original set: while it grants a higher degree of flexibility (since no constraints have to be met) it means that the result of the visualization step entirely depends on the formalization of the mapping procedure and that there is not a direct way to map additional points after obtaining the projections of the data set; this fact renders t-SNE unsuitable for the visualization of streaming data or online scenarios. Additionally, non-parametric techniques are unfitting for large data sets since they display at least a quadratic complexity. To tackle these drawbacks (Gisbrecht et al., 2015) introduces Kernel t-SNE.

Kernel t-SNE is an approach that maintains the flexibility of t-SNE while displaying generalization ability within out-of-sample extensions: it extends the non-parametric t-SNE to an explicit mapping by fixing the parametric form $\mathbf{x} \rightarrow f_w(\mathbf{x}) = \mathbf{y}$ and optimizing the parameters of f_w instead of the projection coordinates. The mapping $f_w = y$ underlying kernel t-SNE follows the form:

$$\mathbf{x} \rightarrow \mathbf{y}(\mathbf{x}) = \sum_j \alpha_j \cdot \frac{k(\mathbf{x}, \mathbf{x}_j)}{\sum_i k(\mathbf{x}, \mathbf{x}_l)} \tag{4}$$

where $a_j \in Y$ are parameters corresponding to points in the projection space and the data x_j are taken as a fixed sample. k is the Gaussian kernel parameterized by the bandwidth σ_j:

$$k(\mathbf{x}, \mathbf{x}_j) = \exp\left(-0.5\|\mathbf{x}-\mathbf{x}_j\|^2/\sigma_j^2\right) \tag{5}$$

This generalized linear mapping allows training to be performed in a simple way (given a set of samples x_i and $y(x_i)$ is available). Parameters

α_j can be determined by a least squares solution of the mapping. Thus, in kernel t-SNE a standard t-SNE is applied to the subset X' to obtain a training set and afterwards the previous analytical solution is used to obtain the parameters of the mapping: once this is done the full set X can be projected in linear time by applying the mapping y. In this work, the bandwidth σ was defined as 0.5 for all experiments, while we used the same range for the perplexity value in kernel t-SNE as the one used in t-SNE.

2.3. Regression model: Support vector regression

After capturing the hyperspectral images and obtaining the reflectance measurements, dimensionality reduction techniques (such as t-SNE) are applied to the data sets to obtain a low-dimensional representation of the points that is better suited to serve as an input for a classification or, for the case of this work, regression model: in this paper we chose the Support Vector Regression (SVR) algorithm (Vapnik, 1999) since it obtained state-of-the-art results (Silva et al., 2018) for the particular application of predicting oenological parameters from hyperspectral images of wine grape berries when compared to Neural Networks (NN) (Janik et al., 2007), Partial Least Squares (PLS) (Arana et al., 2005) or Least-Squares Support Vector Machines (LSSVM) (Cao et al., 2010) models; additionally, to guarantee that for the case studies present in this work the SVR algorithm obtained superior results and that the behaviour exhibited by PCA, t-SNE and Kernel t-SNE followed similar tendencies, we also applied NN for single vintage data sets - results can be found in Appendix A.

The support vector algorithm is described as follows (Smola and Schölkopf, 2004): given a set of training data $\{(x_1,y_1),\ldots,(x_n, y_n)\} \subset \chi \in \mathbb{R}$, where χ represents the space of input patterns, the goal is to determine a function $f(x) = \langle w,x\rangle + b$, $w \in \chi$, $b \in \mathbb{R}$ that deviates a maximum of ε from the real measured targets y_i for the entire training set while, simultaneously, being as flat as possible. This convex optimization problem can be written as the minimization of the Euclidean norm for the cases of a linear function; to extend the SV machine to nonlinear functions the so-called "SV expansion" is introduced:

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)(x_i, x) + b \tag{6}$$

where ω is described as a linear combination of the training patterns and cases with $\alpha_i > 0$ are the support vectors. This SV expansion comes to light by writing the optimization problem in its dual formulation from both the objective function and the corresponding constraints. Additionally, a computationally efficient form of mapping the input vectors into a high-dimensional feature space with a nonlinear mapping comes with the usage of a suitable kernel function, obtaining the nonlinear regression functions of the form:

$$f(x) = \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) \cdot \kappa(x_i, x) + b \tag{7}$$

where κ is the chosen kernel function. This transformation allows the model to solve the optimization problem in a more suitable feature space; there is a wide variety of kernels to choose from that are suited for different types of applications, with local kernels being based on distance (only the data points near each other influence the kernel values) and global kernels being based on dot product (data points far away from each other still influence the kernel values): with the different tests performed in previous works (Silva et al., 2018) we chose a Gaussian kernel for the current paper since it achieved the best results for our application; as for the well-known hyperparameters C and γ, we performed a grid search for each experiment with C values ranging from 80 to 120 and γ ranging from $1e^{-5} - 1e^{-1}$.

2.4. Cross-validation and evaluation metrics

2.4.1. n-Fold Cross-Validation

At this point of the models pipeline it is of paramount importance to guarantee that the results obtained are not skewed in any form by the training stage, since it is very common for these algorithms to suffer from "overfitting" - the statistical model achieving a perfect fit against its training data and being unable to perform accurately against unseen data - to tackle this issue a cross-validation approach was implemented.

The n-fold cross-validation method (Lendasse et al., 2003; Remesan and Mathew, 2015) splits the data set X into K parts of equal size with the kth set forming the validation set X_{val} and the remaining sets forming the training set X_{learn}: the training of a model g is performed using X_{learn} and the error $E_k(g)$ is calculated as:

$$E_k(g) = \frac{\sum_{i=1}^{\frac{N}{K}} \left(g\left(x_i^{val}\right) - y_i^{val}\right)^2}{N/K} \tag{8}$$

with (x_i^{val}, y_i^{val}) the elements of X_{val} and $g(x_i^{val})$ the estimation of y_i^{val} by model g. This process loops for k varying from 1 to K and the average error is computed as:

$$\widehat{E}_{gen}(g) = \frac{\sum_{k=1}^{K} E_k(g)}{K} \tag{9}$$

In this work, we choose the number of folds K varying from a range of 5–10: we choose a smaller number of folds for smaller data sets and a higher number for data sets with more samples.

2.4.2. Trustworthiness and continuity

In order to evaluate the performance of t-SNE and kernel t-SNE and compare it to a standard PCA approach for dimensionality reduction some metric that measures the quality of the low-dimensional embeddings must be employed: based on (Du, 2019; Venna and Kaski, 2006) we implemented the trustworthiness and continuity metrics.

Trustworthiness and Continuity are metrics that attempt to measure the degree of similarity of the local structure of the data between its original high-dimensional state and its low-dimensional visualization obtained via the application of a dimensionality reduction technique. In particular, trustworthiness evaluates if the neighbors chosen are the same in both representations and is defined by:

$$T(k) = 1 - \mathrm{A(k)} \sum_{i=1}^{N} \sum_{j \in U_k(i)} r_{\mathrm{T}}(i,j) - k \tag{10}$$

where N is the sample size, k is the number of nearest neighbors, $A(k)$ is a scaling function and $U_k(i)$ is the set of points that are among the k nearest neighbors of i in the low-dimensional space but not in the high-dimensional one. $r_{\mathrm{T}}(i,j)$ is the rank of point j in $U_k(i)$ according to the pairwise distance from i in the original high-dimensional space. On the other hand, continuity attempts to quantify how well the local structure was maintained after its transformation to a low-dimensional visualization and is defined by:

$$C(k) = 1 - \mathrm{A(k)} \sum_{i=1}^{N} \sum_{j \in V_k(i)} r_C(i,j) - k \tag{11}$$

where $V_k(i)$ is the set of points that are among the k nearest neighbors of data point i in the original high-dimensional space but are not neighbors in the visualization. $r_C(i,j)$ is the rank of the point j in $V_k(i)$ ordered by the pairwise distances between j and i in the low-dimensional visualization.

2.4.3. Root mean square error (RMSE) and determination coefficient R^2

Concerning the evaluation of the performance of the dimensionality reduction techniques we also measure the generalization error of the

regressor trained on the low-dimensional data representation (Sanguinetti, 2008): this process allows the comparison of the prediction results with other state-of-the-art publications. Root mean square error (RMSE) is defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N-1}} \tag{12}$$

where y_i is the reference value and $\hat{y}_i$ is the model estimate. The generalization error is the only well-established practice in literature to evaluate the accuracy of a model and we measure the RMSE for the hold-out test set; however, it seems that for this particular area of research the usage of the determination coefficient (R^2) as an indicator of quality of prediction is very common. R^2 is defined as:

$$R^2 = \left(\frac{\sigma_{y\hat{y}}}{\sigma_y \sigma_y}\right)^2 \tag{13}$$

where $\sigma_{y,\hat{y}}$ is the covariance between y and $\hat{y}$ and σ_i, $\sigma_{\hat{y}}$ are the respective standard deviations. While we present the R^2 indicator in the results section to serve as a sort of comparison benchmark to other works in literature for the prediction of oenological parameters in hyperspectral images of wine grape berries, it is important to denote that, according to (Spiess and Neumeyer, 2010), R^2 is not well-defined for non-linear regression since it shows extreme bias to higher parameterized models and, in a background of low and medium experimental noise, this indicator cannot compensate the effect of the number of increasing parameters.

2.5. Grape sampling and data set analysis

Regarding the variety of wine grape berries used to frame the data sets for this work, we chose grapes of the native Portuguese variety Touriga Franca (TF) mainly due to its' importance for Port wine production in the Portuguese Douro region, as well as due to its' importance to our industrial partner Symington Family Estates: the TF variety is so important for both producers and oenologists due to its' resistance to most plant diseases, the strong flavour and aroma of the resulting wine and due to its' high concentration of tannins, which guarantees that the wine will age well. The grapes were harvested from Quinta do Bonfim in Pinhão, Portugal in the years (refers to the years of sample collecting, do not mistake for year of vine tree plantation) of 2012 (240 samples), 2013 (81 samples), 2014 (120 samples), 2016 (407 samples) and 2017 (540 samples) from different regions in the vineyard and with different levels of maturity; each sample is composed of six grape berries collected from a single bunch and the ground-truth results are obtained via laboratory analysis using validated standard methods (Carbonneau and Champagnol, 1993; Office International de la Vigne and du Vin, 1990).

Henceforth, we conducted a study of the models ability to predict the pH index and sugar content on different vintages of wine grape berries from the TF variety: we chose to evaluate these oenological parameters since they are highly researched due to their correlation with flavour, colour and overall grape ripeness stage; additionally, due to several factors such as climate change, soil quality, sun exposition, water assessment, altitude and harvest time, a large variability is present in the vineyards and consequently in the quality of the grapes and their oenological profile, which makes the evaluation of the models' capacity across different vintages a true test to its' generalization potential. Appendix B contains a descriptive statistics analysis of the ground-truth results for each oenological parameter in the different TF data sets used; 7 presents ANOVA (one-way analysis of variance) tests to verify significant differences between the means of the different data sets.

In order to achieve a more comprehensible form of presenting the results, we decided to split the analysis according to the characteristics of the training set and the hold-out test set for each run:

Table 1
Experiment details for each case study.

Case	Training/Validation set	Test set
A1	TF 2012	TF 2012
A2	TF 2013	TF 2013
A3	TF 2014	TF 2014
B1	TF 2012/13	TF 2012/13
B2	TF 2012/13/14	TF 2012/13/14
B3	TF 2012/13/14/16	TF 2012/13/14/16
C1	TF 2012	TF 2013
C2	TF 2012/13	TF 2014
C3	TF 2012/13/14/16	TF 2017

- Case A*n*: cases where the training and test sets use the same single vintage of TF wine grape berries;
- Case B*n*: cases where the training set employs multiple vintages of TF wine grape berries and the hold-out test set uses the same multiple vintages of TF wine grape berries.
- Case C*n*: cases where the training set uses single/multiple vintages of TF wine grape berries and the hold-out test set uses a different vintage of TF wine grape berries.

For each case study we create a hold-out test set with 10% of the size of the correspondent training set; for each training set we perform cross-validation with a 90/10% split into training/validation folders. Table 1 provides details on all the experiments performed.

In each of the experiments previously described we present the RMSE and R^2 obtained by the regressor for the hold-out test set for each oenological parameter (pH index and sugar content) and for each dimensionality reduction method employed (PCA, t-SNE and kernel t-SNE) - as mentioned previously, we compare the performance of t-SNE and kernel t-SNE against PCA since in our previous work (Silva and Melo-Pinto, 2021), for these same data sets and case studies, PCA outperformed a wide variety of nonlinear methods establishing itself as the best technique to reduce our inputs, which goes in accordance to several studies in literature that show that, for real-world data, PCA still outperforms the most advanced techniques introduced in recent years; the trustworthiness and continuity metrics are calculated for each data set for every dimensionality reduction method implemented. Appendix D summarizes the best published results in literature and how they compare with the best results in this work, split by corresponding case studies.

3. Results and discussion

3.1. Analysis

Table 2 and Table 3 present the results (best results are underlined) obtained for the trustworthiness and continuity measures for all data sets. Observing these tables, it is noticeable that t-SNE achieves the best results in trustworthiness across the different data sets but the continuity results are a bit closer: this means that while t-SNE does a better job of choosing the same neighbors in both the higher and lower dimensional representations, how well the local structure was maintained after its transformation (defined by continuity) is on par between both methods; however, it is important to denote that for the data

Table 2
Trustworthiness $T(20)$ for each dimensionality reduction technique in each individual data set.

Trustworthiness					
Method	TF 2012	TF 2013	TF 2014	TF 2016	TF 2017
PCA	0.9949	0.9935	0.9972	0.9832	0.9702
t-SNE	0.9966	0.9947	0.9935	0.9975	0.9981
Kernel t-SNE	0.9871	0.9761	0.9649	0.9840	0.9866

Table 3
Continuity $C(20)$ for each dimensionality reduction technique in each individual data set.

Continuity					
Method	TF 2012	TF 2013	TF 2014	TF 2016	TF 2017
PCA	0.9965	0.9944	0.9978	0.9912	0.9866
t-SNE	0.9966	0.9942	0.9910	0.9954	0.9926
Kernel t-SNE	0.9857	0.9539	0.9571	0.9852	0.9773

sets with a higher number of samples and a higher intrinsic dimensionality, t-SNE gets the best results which might be an indicator of this method being better suited to reduce the dimensionality of more complex samples, a trait that is also present on Kernel t-SNE results.

Table 4 showcases the results obtained in the prediction of the pH index for the different case studies and dimensionality reduction methods. In a direct comparison, the t-SNE technique gets the best results for almost every single case study, with smaller error rates and higher determination coefficient values. As highlighted previously by the trustworthiness and continuity measures, it is also important to notice that the Kernel t-SNE method achieves smaller error rates in the last case studies in comparison to the PCA method: these are the cases where training sample number increases significantly and where the intrinsic dimensionality of the data rises. While the regressor performance is quite stable across the different case studies, it is still noticeable that the smallest error rates are obtained in cases where the training and test sets use the same single vintage of TF wine grape berries: this shows that despite having a very strong generalization capacity, the model still cannot predict the pH index for cases of unseen vintages in the testing phase with such certainty as it does for when the training set includes samples of that same vintage: this can be considered normal, since the complexity of the prediction significantly increases (Appendix C Table C.1 shows that there is a significant difference in the means between almost every single vintage) and an even bigger drop-off in results would not be surprising, since we are using a shallow learning regressor to operate the predictions and pH index values are extremely vulnerable to even the smallest changes in external factors (such as weather, water availability, etc.). Additionally, we conducted a paired *t*-test between the predictions obtained by each of the dimensionality reduction methods and no difference in the means between the sets was found.

Table 5 presents the results obtained for the prediction of sugar content for the different case studies and dimensionality reduction methods: once again, the t-SNE technique achieves the best results for almost every case study, with smaller error rates and higher determination coefficient values; additionally, it is once again important to emphasize that both t-SNE and Kernel t-SNE results for cases where the intrinsic dimensionality and number of samples rises are better than the ones obtained by PCA. As for the generalization capacity of the regressor, for the sugar content prediction the drop-off in results (higher error rate) is more noticeable than for the case of the pH index, which could be expected since we are working with more sparsely distributed sample points (Appendix B presents descriptive statistics of the samples for each data set in Table B.1 and Table B.2); this increase in error rate can be even more perceptible in cases where the TF 2014 vintage is employed in training or testing (cases A3, B2, B3, C2 and C3), which can be explained by the fact that this particular vintage had particularly low values for the sugar content, with a much smaller mean value when compared to the other vintages, causing the regressor to have bigger problems in generalizing its predictions - however, we still consider the generalization capacity of the model to be very acceptable, specially considering we are using a shallow learning regressor that lacks the capability of more powerful deep learning techniques. The paired t-test to investigate statistical significant differences between the means of the predictions given by each dimensionality reduction method showed that there were significant differences in the means of the predictions obtained by the PCA and Kernel t-SNE methods when compared to the PCA for case studies A3 and C3, which possibly justifies the difference in the results with the fact that the regressor highlighted different features for learning in these sets.

Table 4
Results obtained for pH index in the hold-out test set for each case study.

pH Index						
Case	PCA		t-SNE		Kernel t-SNE	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
A1	0.786	0.186	0.870	0.148	0.825	0.177
A2	0.810	0.199	0.947	0.106	0.899	0.139
A3	0.753	0.139	0.829	0.120	0.788	0.148
B1	0.814	0.171	0.876	0.158	0.813	0.170
B2	0.818	0.163	0.863	0.171	0.803	0.168
B3	0.783	0.174	0.832	0.162	0.757	0.185
C1	0.845	0.229	0.953	0.153	0.845	0.189
C2	0.818	0.195	0.808	0.152	0.738	0.161
C3	0.797	0.245	0.869	0.120	0.860	0.133

Table 5
Results obtained for sugar content in the hold-out test set for each case study.

Sugar Content (°Brix)						
Case	PCA		t-SNE		Kernel t-SNE	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
A1	0.922	1.146	0.949	0.955	0.926	1.113
A2	0.940	1.470	0.985	0.737	0.982	0.950
A3	0.898	2.155	0.958	1.279	0.817	1.288
B1	0.935	1.791	0.944	1.180	0.918	1.492
B2	0.916	1.601	0.907	1.545	0.851	1.593
B3	0.876	1.304	0.890	1.430	0.866	1.746
C1	0.940	1.544	0.964	1.510	0.951	1.726
C2	0.839	3.558	0.940	3.789	0.868	3.559
C3	0.830	2.964	0.840	1.863	0.809	1.776

3.2. Comparison

Comparing our results with top state-of-art works (Appendix D Table D.1 provides a comparative review table), t-SNE obtained the best RMSE and R^2 values for sugar content estimation out of all studies where the training and test sets use the same single vintage of wine grape berries and also out of all studies where the test set uses a different vintage of wine grape berries; regarding case study B (same multiple vintages of wine grape berries employed in both the training and hold-out test sets) t-SNE achieved on par and competitive results, only being outdone by the Convolutional Neural Networks (CNN) model in (Gomes et al., 2021a) and by the Partial Least Squares (PLS) models presented by (Caballero et al., 2011) and (Fadock et al., 2016): in the authors opinion, this emphasizes even more the capabilities of the t-SNE technique for dimensionality reduction, since we are able to obtain competitive or superior results with a shallow learning regressor versus models that already employ more advanced, deep learning solutions for prediction; considering the differences in methodology between different research groups (different data sets, different setups for hyperspectral imaging, different modes for reflectance/transmittance/interactance measurements, different techniques for prediction, etc.) we also consider that a direct comparison, with the same pipeline and samples, between different techniques provide better conclusions about the dimensionality reduction step and, in this matter, we showcase that t-SNE achieves better results than PCA for the same case studies, which is a novelty since no other works can be found that report this superiority in real-world data.

As for the pH index estimation, t-SNE achieved very competitive results, even obtaining the best RMSE and R^2 values of all studies where the training and test sets use the same single vintage of wine grape berries; when comparing case studies with multiple vintages of wine

grape berries or with a different vintage used in the hold-out test set, despite showcasing a very strong performance, our model still falls a bit short out of the results obtained by the CNN in (Gomes et al., 2021a, 2021b): as mentioned previously, we believe that a direct comparison between techniques provides more meaningful information regarding the dimensionality reduction step, since it is hard to pinpoint if the difference in results derives from a weaker dimensionality reduction technique, from a weaker regressor, or from the overall differences in methodology and samples - in this matter and in our case studies, we once again highlight the superiority of t-SNE when compared to PCA, independently of the regressor employed (Appendix A shows that while using a NN regressor the t-SNE still outperforms PCA).

4. Conclusions

An evaluation of the performance of t-SNE and Kernel t-SNE techniques for dimensionality reduction was carried out in real-world data of hyperspectral images of wine grape berries. Our study combined these techniques with a machine learning regressor (Support Vector Regression) to predict sugar content and pH index values in different vintages of wine grape berries, a highly complex problem due to high variability in samples across different vintages and varieties of wine grapes. Our results show that not only t-SNE is a technique suited for dimensionality reduction tasks (it is normally used for visualization purposes only), but it even surpasses the results obtained by PCA for almost every single case study, a very rare achievement since most recent dimensionality reduction techniques introduced showcase a strong performance on artificial data sets but fail to replicate it in real-world data, usually falling short of the performance obtained by PCA. Additionally, Kernel t-SNE also showed a strong performance, opening the possibility of using a t-SNE extension for dimensionality reduction tasks in streaming data or online scenarios.

Comparing our results with state-of-art works, our model combined of a shallow learning regressor with t-SNE to reduce the dimensionality of the inputs achieved either better or on par results with the ones presented in literature, even competing with works using more refined deep learning models (like CNNs); this highlights the capability of t-SNE, since previous models that combined a shallow learning regressor with PCA could not compete with the results obtained by more advanced models. Additionally, the development of accurate shallow learning models for oenological parameter estimation from hyperspectral images of wine grapes is extremely important, since acquiring the necessary number of samples to allow for the usage of a deep learning model that does not overfit to the training set and has generalization capacity is very costly and time-consuming.

Future works should expand on the possibility of using t-SNE in the dimensionality reduction step when combined with deep learning models, since it is possible that an increase in performance could be achieved: additionally, we also believe that further work in this area should be directed into deep learning models and artificial data synthesis, since it is very difficult and costly to obtain new samples and there is a huge variability present in the spectra of different varieties and vintages of wine grapes.

CRediT authorship contribution statement

Rui Silva: Conceptualization, Software, Validation, Formal analysis, Investigation, Writing – original draft, Visualization. **Pedro Melo-Pinto:** Methodology, Resources, Data curation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by National Funds by FCT — Portuguese Foundation for Science and Technology, under the project UIDB/04033/2020. The authors also gratefully acknowledge the support from National funding by FCT, Portuguese Foundation for Science and Technology, through the individual research grant (SFRH/BD/137216/2018) and from NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research.

Appendix A

Table A.1
Results obtained for pH index in the hold-out test set for case studies A with SVR and NN.

		pH index					
Model	Case	PCA		t-SNE		Kernel t-SNE	
		R^2	RMSE	R^2	RMSE	R^2	RMSE
SVR	A1	0.786	0.186	0.870	0.148	0.825	0.177
	A2	0.810	0.199	0.947	0.106	0.899	0.139
	A3	0.753	0.139	0.829	0.120	0.788	0.148
NN	A1	0.764	0.202	0.830	0.168	0.790	0.197
	A2	0.787	0.496	0.870	0.187	0.865	0.231
	A3	0.738	0.176	0.802	0.120	0.761	0.150

Table A.2
Results obtained for sugar content in the hold-out test set for case studies A with SVR and NN.

Sugar content							
Model	Case	PCA		t-SNE		Kernel t-SNE	
		R^2	RMSE	R^2	RMSE	R^2	RMSE
SVR	A1	0.922	1.146	0.949	0.955	0.926	1.113
	A2	0.940	1.470	0.985	0.737	0.982	0.950
	A3	0.898	2.155	0.958	1.279	0.817	1.288
NN	A1	0.926	1.695	0.928	1.036	0.904	1.404
	A2	0.936	1.641	0.981	1.392	0.947	1.452
	A3	0.897	2.220	0.904	1.876	0.798	2.287

Appendix B

Table B.1
Descriptive statistics for the pH index of all TF vintages.

pH index								
Variety	N	Mean	95% CI	St.Dev.	95% CI	Min	Median	Max
TF 2012	240	3.55	(3.51; 3.60)	0.35	(0.32; 0.38)	2.85	3.58	4.23
TF 2013	81	3.72	(3.64; 3.80)	0.35	(0.31; 0.42)	3.05	3.74	4.44
TF 2014	120	3.49	(3.45; 3.54)	0.26	(0.23; 0.30)	2.93	3.51	3.97
TF 2016	404	3.85	(3.83; 3.88)	0.28	(0.26; 0.30)	3.09	3.84	4.60
TF 2017	538	3.90	(3.88; 3.92)	0.19	(0.18; 0.20)	3.29	3.92	4.97

Table B.2
Descriptive statistics for the sugar content of all TF vintages.

Sugar content (°Brix)								
Variety	N	Mean	95% CI	St.Dev.	95% CI	Min	Median	Max
TF 2012	240	16.93	(16.50; 17.35)	3.34	(3.07; 3.67)	9.06	17.06	24.71
TF 2013	82	19.45	(18.66; 20.23)	3.58	(3.11; 4.24)	8.10	20.03	25.00
TF 2014	120	13.56	(12.89; 14.21)	3.66	(3.25; 4.19)	7.87	13.00	25.66
TF 2016	404	17.83	(17.57; 18.09)	2.62	(2.45; 2.82)	10.07	17.54	26.03
TF 2017	538	19.86	(19.57; 20.14)	3.37	(3.18; 3.58)	10.94	20.34	30.08

Appendix C

Table C.1
One-way ANOVA for the pH index of the laboratory results.

Tukey simultaneous tests for differences of means					
pH index					
Diff. of Levels	Diff. of Means	SE. of Diff	95% CI	T-Value	ρ-value
TF2013 - TF2012	0.17	0.03	(0.07; 0.26)	4.85	0.000
TF2014 - TF2012	−0.06	0.03	(−0.14; 0.02)	−1.99	0.272
TF2016 - TF2012	0.30	0.02	(0.24; 0.37)	13.92	0.000
TF2017 - TF2012	0.35	0.02	(0.29; 0.40)	16.84	0.000
TF2014 - TF2013	−0.22	0.04	(−0.33; −0.12)	−5.88	0.000
TF2016 - TF2013	0.14	0.03	(0.05; 0.22)	4.20	0.000
TF2017 - TF2013	0.18	0.03	(0.10; 0.27)	5.74	0.000
TF2016 - TF2014	0.36	0.03	(0.29; 0.44)	13.05	0.000
TF2017 - TF2014	0.40	0.03	(0.33; 0.48)	15.15	0.000
TF2017 - TF2016	0.05	0.02	(0.00; 0.09)	2.63	0.065

H_0: All means are equal; Significance: $\alpha = 0.05$; Individual Confidence = 99.55%.

Table C.2
One-way ANOVA for the sugar content of the laboratory results.

Tukey simultaneous tests for differences of means					
Sugar content (°Brix)					
Diff. of levels	Diff. of means	SE. of Diff	95% CI	T-value	ρ-value
TF2013 - TF2012	2.52	0.41	(1.40; 3.64)	6.15	0.000
TF2014 - TF2012	−3.37	0.35	(−4.35; −2.40)	−9.41	0.000
TF2016 - TF2012	0.90	0.26	(0.19; 1.62)	3.46	0.005
TF2017 - TF2012	2.93	0.25	(2.25; 3.61)	11.77	0.000
TF2014 - TF2013	−5.89	0.46	(−7.15; −4.64)	−12.83	0.000
TF2016 - TF2013	−1.62	0.39	(−2.68; −0.58)	−4.16	0.000
TF2017 - TF2013	0.41	0.38	(−0.63; 1.45)	1.08	0.819
TF2016 - TF2014	4.28	0.33	(3.37; 5.19)	12.83	0.000
TF2017 - TF2014	6.30	0.32	(5.42; 7.19)	19.47	0.000
TF2017 - TF2016	2.03	0.21	(1.45; 2.60)	9.60	0.000

H_0: All means are equal; Significance: $\alpha = 0.05$; Individual Confidence = 99.55%.

Appendix D

Table D.1
Summary of the best results published in literature for the prediction of oenological parameters on hyperspectral images of wine grape berries.

			RMSE		R^2	
Cases	Authors	Model	Sugar (°Brix)	pH Index	Sugar	pH Index
A			**0.737**	**0.106**	0.985	**0.947**
B	PW	SVR	1.180	0.158	0.944	**0.876**
C			1.510	**0.153**	0.964	**0.953**
	(Arana et al., 2005)	PLS	1.270	–	0.710	–
	(Bueno et al., 2014)	MPLS	1.370	0.120	0.990	0.940
	(Cao et al., 2010)	LSSVM	0.960	–	0.907	–
	(Cao et al., 2010)	PLS	0.925	–	0.914	–
	(Costa et al., 2019)	PLS	1.150	–	0.950	–
A	(Cozzolino et al., 2004)	MPLS	–	0.150	–	0.850
	(Fernandes et al., 2015)	NN	0.950	0.180	0.920	0.730
	(Gomes et al., 2017a)	NN	0.955	–	0.924	–
	(Gomes et al., 2017a)	PLS	0.939	–	0.929	–
	(Silva et al., 2018)	SVR	0.804	0.142	0.964	0.887
	(Silva and Melo-Pinto, 2021)	SVR	1.100	–	0.930	–
	(Caballero et al., 2011)	MPLS	1.000	0.120	0.910	0.870
	(Fadock et al., 2016)	PLS	1.090	0.060	0.700	0.720
	(Gomes et al., 2017b)	NN	–	0.191	–	0.723
	(Gomes et al., 2017a)	NN	1.355	–	0.917	–
B	(Gomes et al., 2017a)	PLS	1.344	–	0.948	–
	(Gomes et al., 2021a)	CNN	0.755	0.110	–	–
	(Gomes et al., 2021b)	CNN	0.970	–	0.920	–
	(Silva et al., 2018)	SVR	1.048	0.170	0.948	0.830
	(Silva and Melo-Pinto, 2021)	SVR	0.922	–	0.969	–
C	(Gomes et al., 2021a)	CNN	1.085	0.183	–	–
	(Silva and Melo-Pinto, 2021)	NN	1.060	–	0.986	–

PW: Present Work.

References

Alibert, Y., 2019. New metric to quantify the similarity between planetary systems: application to dimensionality reduction using t-SNE. Astron. Astrophys. 624, A45.

Anowar, F., Sadaoui, S., Selim, B., 2021. Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). Comput. Sci. Rev. 40, 100378.

Arana, I., Jarén, C., Arazuri, S., 2005. Maturity, variety and origin determination in white grapes (vitis vinifera l.) using near infrared reflectance technology. J. Near Infrared Spectrosc. 13, 349–357.

Bueno, J., Hernández-Hierro, J., Rodrguez-Pulido, F., Heredia, F., 2014. Determination of technological maturity of grapes and total phenolic compounds of grape skins in red and white cultivars during ripening by near infrared hyperspectral image: a preliminary approach. Food Chem. 152, 586–591.

Caballero, V., Pérez-Marn, D., López, M., Sánchez, M., 2011. Optimization of nir spectral data management for quality control of grape bunches during on-vine ripening. Sensors 11, 6109–6124.

Cao, F., Wu, D., He, Y., 2010. Soluble solids content and ph prediction and varieties discrimination of grapes based on visible–near infrared spectroscopy. Comput. Electron. Agric. 71, S15–S18.

Carbonneau, A., Champagnol, F., 1993. Nouveaux systèmes de culture integré du vignoble. Programme AIR.

Costa, D., Mesa, N., Freire, M., Ramos, R., Mederos, B., 2019. Development of predictive models for quality and maturation stage attributes of wine grapes using Vis-nir reflectance spectroscopy. Postharvest Biol. Technol. 150, 166–178.

Cozzolino, D., Cynkar, W., Janik, L., Dambergs, B., Francis, L., Gishen, M., 2004. Measurement of colour, total soluble solids and ph in whole red grapes using visible and near infrared spectroscopy. Proceedings of the 12th Australian Wine Industry Technical Conference, Melbourne, Australia, pp. 24–29.

Devassy, B., George, S., 2020. Dimensionality reduction and visualisation of hyperspectral ink data using t-SNE. Forensic Sci. Int. 311, 110194.

Du, T., 2019. Dimensionality reduction techniques for visualizing morphometric data: comparing principal component analysis to nonlinear methods. Evol. Biol. 46, 106–121.

Fadock, M., Brown, R., Reynolds, A., 2016. Visible-near infrared reflectance spectroscopy for nondestructive analysis of red wine grapes. Am. J. Enol. Vitic. 67, 38–46.

Fernandes, A., Oliveira, P., Moura, J., Oliveira, A., Falco, V., Correia, M., Melo-Pinto, P., 2011. Determination of anthocyanin concentration in whole grape skins using hyperspectral imaging and adaptive boosting neural networks. J. Food Eng. 105, 216–226.

Fernandes, A., Franco, C., Mendes-Ferreira, A., Mendes-Faia, A., da Costa, P., Melo-Pinto, P., 2015. Brix, pH and anthocyanin content determination in whole port wine grape berries by hyperspectral imaging and neural networks. Comput. Electron. Agric. 115, 88–96.

Gao, L., Gu, D., Zhuang, L., Ren, J., Yang, D., Zhang, B., 2019. Combining t-distributed stochastic neighbor embedding with convolutional neural networks for hyperspectral image classification. IEEE Geosci. Remote Sens. Lett. 17, 1368–1372.

Gisbrecht, A., Mokbel, B., Hammer, B., 2012. Linear basis-function t-SNE for fast nonlinear dimensionality reduction. The 2012 International Joint Conference on Neural Networks, pp. 1–8.

Gisbrecht, A., Schulz, A., Hammer, B., 2015. Parametric nonlinear dimensionality reduction using kernel t-SNE. Neurocomputing 147, 71–82.

Gomes, V., Melo-Pinto, P., 2021. Towards robust Machine Learning models for grape ripeness assessment. IEEE 18th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 1–5.

Gomes, V., Fernandes, A., Faia, A., Melo-Pinto, P., 2014a. Comparison of different approaches for the Prediction of Sugar Content in Whole Port Wine Grape Berries using Hyperspectral Imaging. ENBIS 14: 14th Annual Conference of the European Network for Business and Industrial Statistics, p. 1.

Gomes, V., Fernandes, A., Faia, A., Melo-Pinto, P., 2014b. Determination of Sugar Content in Whole Port Wine Grape Berries Combining Hyperspectral Imaging with Neural Networks Methodologies. IEEE Symposium Series on Computational Intelligence.

Gomes, V., Fernandes, A., Faia, A., Melo-Pinto, P., 2017a. Comparison of different approaches for the prediction of sugar content in new vintages of whole port wine grape berries. Comput. Electron. Agric. 140, 244–254.

Gomes, V., Fernandes, A., Martins-Lopes, P., Pereira, L., Faia, A., Melo-Pinto, P., 2017b. Characterization of neural network generalization in the determination of pH and anthocyanin content of wine grape in new vintages and varieties. Food Chem. 218, 40–46.

Gomes, V., Mendes-Ferreira, A., Melo-Pinto, P., 2021a. Application of hyperspectral imaging and deep learning for robust prediction of sugar and pH levels in wine grape berries. Sensors 21, 3459.

Gomes, V., Reis, M., Rovira-Más, F., Mendes-Ferreira, A., Melo-Pinto, P., 2021b. Prediction of sugar content in port wine vintage grapes using machine learning and hyperspectral imaging. Processes 9, 1241.

Gowen, A., O'Donnell, C., Cullen, P., Downed, G., Frias, J., 2007. Hyperspectral imaging – an emerging process analytical tool for food quality and safety control. Trends Food Sci. Technol. 18, 590–598.

Hall, A., Lamb, D., Holzapfel, B., Louis, J., 2002. Optical remote sensing applications in viticulture - a review. Aust. J. Grape Wine Res. 36–47.

Hariharan, S., 2021. Analysing effect of t-SNE and 1-D CNN on performance of hyperspectral image classification. Turk. J. Comput. Math. Educ. 12, 1828–1833.

Hinton, G., Roweis, S., 2002. Stochastic neighbor embedding. Advances in Neural Information Processing Systems, pp. 833–840.

Janik, L., Cozzolino, D., Dambergs, R., Cynkar, W., Gishen, M., 2007. The prediction of total anthocyanin concentration in red-grape homogenates using visible-near-infrared spectroscopy and artificial neural networks. Anal. Chim. Acta 594, 107–118.

Lendasse, A., Wertz, V., Verleysen, M., 2003. Model selection with cross-validations and bootstraps—Application to time series prediction with rbfn models. Artificial Neural

Networks and Neural Information Processing—ICANN/ICONIP 2003. Springer, pp. 573–580.
Miao, A., Zhuang, J., Tang, Y., He, Y., Chu, X., Luo, S., 2018. Hyperspectral image-based variety classification of waxy maize seeds by the t-SNE model and procrustes analysis. Sensors 18, 4391.
Office International de la Vigne and du Vin, 1990. Recueil des méthodes internationales d'analyse des vins et des moûts. OIV.
Pouyet, E., Rohani, N., Katsaggelos, A., Cossairt, O., Walton, M., 2018. Innovative data reduction and visualisation stratagy for hyperspectral imaging datasets using t-SNE approach. Pure Appl. Chem. 90, 493–506.
Remesan, R., Mathew, J., 2015. Model data selection and data pre-processing approaches. Hydrological Data Driven Modelling. Springer, pp. 41–70.
Sanguinetti, G., 2008. Dimensionality reduction of clustered data sets. IEEE Trans. Pattern Anal. Mach. Intell. 30, 535–540.
Schulz, A., Hammer, B., 2015. Discriminative dimensionality reduction for regression problems using the fisher metric. International Joint Conference on Neural Networks, pp. 1–8.
Silva, R., Melo-Pinto, P., 2021. A review of different dimensionality reduction methods for the prediction of sugar content from hyperspectral images of wine grape berries. Appl. Soft Comput. 113.
Silva, R., Gomes, V., Mendes-Faia, A., Melo-Pinto, P., 2018. Using support vector regression and hyperspectral imaging for the prediction of oenological parameters on different vintages and varieties of wine grape berries. Remote Sens. 10, 312.
Smola, A., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 199–222.
Spiess, A., Neumeyer, N., 2010. An evaluation of r2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. BMC Pharmacol. 10, 1–11.
Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9.
Vapnik, V., 1999. The Nature of Statistical Learning Theory. Springer Science & Business Media.
Venna, J., Kaski, S., 2006. Visualizing Gene Interaction Graphs with Local Multidimensional Scaling. ESANN, Citeseer, pp. 557–562.
Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemom. Intell. Lab. Syst. 2, 37–52.
Zhang, J., Chen, L., Zhou, L., Liang, X., Li, J., 2018. An efficient hyperspectral image retrieval method: deep spectral-spatial feature extraction with DCGAN and dimensionality reduction using t-SNE-based NM hashing. Remote Sens. 10, 271.