IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 September 2023, accepted 14 October 2023, date of publication 18 October 2023, date of current version 24 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3325676

RESEARCH ARTICLE

Determining the Importance of Physicochemical Properties in the Perceived Quality of Wines

AMALIA LUQUE[1,2], (Member, IEEE), MIRKO MAZZOLENI[2], (Member, IEEE), FRANCISCO ZAMORA-POLO[1], ANTONIO FERRAMOSCA[2], (Member, IEEE), JUAN RAMÓN LAMA[1], AND FABIO PREVIDI[2], (Member, IEEE)

[1]Departamento de Ingeniería del Diseño, Escuela Politécnica Superior, Universidad de Sevilla, 41011 Seville, Spain
[2]Department of Management, Information and Production Engineering, University of Bergamo, Dalmine, 24044 Bergamo, Italy

Corresponding author: Amalia Luque (amalialuque@us.es)

This work was supported in part by the GOYA—Antonio Unanue in Engineering in the Agrifood Industry Chair of the University of Seville, Spain; and in part by the VII Own Research and Transfer Plan 2023 of the University of Seville under Project 2023/00000378.

ABSTRACT Wine is a relevant part of the diet in many countries, showing significant nutritional properties, providing health benefits to consumers, and having a significant weight in economy. Also, wine plays an important role in many cultures as a part of their social relationships, feasts, or religion where some of them may become a sign of luxury and distinction. For those reasons, objective and subjective quality of wines is an important issue in their production and marketing. To improve wine excellence, some production methods try to relate its physicochemical properties to the quality as it is perceived by humans. Then, modern data prescriptive analysis can be applied to measure the importance (the influence) of each wine attribute. This paper examines and compare several metrics of the attribute importance and its application to the quality-aware design and production of wines. Moreover, for the cases where the perceived quality is characterized using a discrete value, a novel importance metric, based on the Jensen-Shannon Divergence (JSD) is introduced and compared to the existing ones. The results show that JSD clearly overperforms other metrics previously proposed in the literature. Also, it can be asserted that JSD properly reflects the importance of discrete multivalued functions. The results, using this metric in an importance performance analysis of a public wine dataset, show that the main physicochemical attributes of a red wine are citric acidity, alcohol, sulphates and fixed acidity. As for the white wine case, the main attributes are alcohol, free sulfure dioxide and pH.

INDEX TERMS Importance metric, importance performance analysis, Jensen Shannon divergence, wine quality.

I. INTRODUCTION

Wine production and consumption have a significant weight in food (and general) economy [1]. This is particularly true for Mediterranean countries like Italy [2], Spain [3] and other European Union states [4]. Wine is undoubtedly a relevant part of the diet in many countries, having significant nutritional properties [5] and being able to provide positive health benefits [6], [7].

Also, wine plays an important role in many cultures [8] as a part of their social relationships, feasts or religion. Wine is appreciated not only by its euphoric properties but also some of them may become a sign of luxury and distinction [9]. For those reasons, objective and subjective quality of wines is an important issue in their production and marketing [10], [11].

Quality perception of the wine drinking experience is a multidimensional issue where the level of consumers' expertise and involvement must be considered [12]. This quality can be obtained using implicit measures, such as electroencephalography (EEG), skin conductance, heart rate, eyes movements, or speed of drinking [13]. However, most of the quality assessment of wines (and other food products) are based on explicit methods, that is, on verbal self-reported

The associate editor coordinating the review of this manuscript and approving it for publication was Xiao-Sheng Si.

measurements that ask participants to report their feelings and emotions about consumption [14].

To improve wine excellence, some production methods try to relate its physicochemical properties to the quality as it is perceived by humans [15], [16]. Then, modern data analytics methods can be employed to description, diagnostic and prediction of the quality of wines [17], [18]. But for product design, the most interesting data analytic methods involve some kind of prescriptive assessment, indicating what should be done to improve the quality of the products [19]. A comprehensive conceptual framework for the optimization of environmental sustainability in engineering projects, both for products and industrial facilities or processes is described in [20].

One key element in prescriptive analysis is to measure the importance (the influence) that a certain physicochemical attribute has on the perceived quality of the wine [21]. The degree of influence of a certain driving attribute on a target variable is commonly measured using linear regression or Pearson's correlation coefficients [22]. They work quite well on linear relationships between variables but fail on more complex functions [23]. To address those cases (nonlinear, non-monotonic or multivalued relationships), other authors have proposed more advanced metrics that will be further described.

This paper examines and compares those metrics and their application to the quality-aware design and production of wines. Moreover, for the cases where the perceived quality is characterized using a discrete value, a novel importance metric, based on the Jensen-Shannon Divergence (JSD) is introduced and compared to the existing ones [24].

Finally, for the prescriptive analysis, that is, to suggest the physicochemical properties of the wine that should be specially considered, an Importance Performance Analysis (IPA) is proposed [25], [26]. Then, applying an IPA with a JSD-based importance, the most quality-driving attributes of wines are identified in two public datasets.

This prescriptive analysis is carried out using a data-driven approach where statistical methods are applied to unveil the strength of the relationship between each objective physicochemical property and the subjective perceived quality. But, like any data-driven method, it works like a black box that does not report of the causes of these relationships as there is no underlying causal model.

Although the research is centered around wine production, their methodologies and conclusions can easily be extended to many other fields and applications. For instance, in the electric and electronic disciplines, it could be applied to the characterization of a semiconductor using specialized equipment which shows an analogous relationship between the multiple factors that define the measuring process, and the value obtained for the electrical property [27], [28], [29], [30].

The practical application of this research is dual. First, it has a straightforward impact on the winemaking process by identifying the physical properties with the greatest influence on the perceived quality and, therefore, those subject to their preferential optimization. And second, the methodology and the metric derived for measuring the importance of each physical property can be easily extended to the production of many other food and beverage products and can even be applied to much broader engineering fields.

The body of the paper starts in section II, where the wine dataset and the methodology employed through the paper are formalized. The methodology description includes the concept of importance, how this concept is implemented using simple metrics (based on linear regression and correlation) and how it should be interpreted in the cases of multivalued functions. It also describes a set of synthetic datasets and defines performance metrics to compare different definitions of importance.

The remaining of the paper is organized as follows: in section III, different importance metrics are detailed and applied to the synthetic datasets; then, the main results obtained applying these techniques are summarized in section IV; and finally, these results are discussed in section V, presenting the main conclusions in section VI.

II. MATERIAL AND METHODS

A. WINE DATASETS

For the application of the ideas presented in this paper, one of the most popular datasets by the University of California Irvine (UCI) Machine Learning Repository [31] will be used. In particular, we will focus on the Wine Quality Data Set [32], containing two datasets related to red and white *vinho verde* wine samples, from the north of Portugal. The first dataset contains 1599 samples of different red wines, while the second dataset includes 4898 samples of white wines. Each sample is characterized by 11 physicochemical features (see Table 1) and a quality score (q) from 0 (very bad) to 10 (excellent), obtained as the median value of the blind evaluations of, at least, 3 sensory assessors. It must be noted that the quality score (q) is an integer number. A discussion on the required number of samples for validation purposes can be found in [33].

For the red wine dataset, the values of each feature for the samples with the same quality are distributed according to the estimated probability density functions (pdfs) depicted in Fig.1. A similar result can be obtained for the white wine dataset.

In these plots a first insight of the influence of each physicochemical feature in the perceived quality can be obtained. Those features presenting more separated pdfs will probably have a higher impact on the quality. This intuition will be justified through the paper.

A more direct way to represent the relationship between the values of a certain feature and the perceived quality would be by drawing each sample as a dot in the feature-quality plane. But as there are more than one thousand samples and the quality has only integer values, these plots would show a high density of dots which would hide the pursued

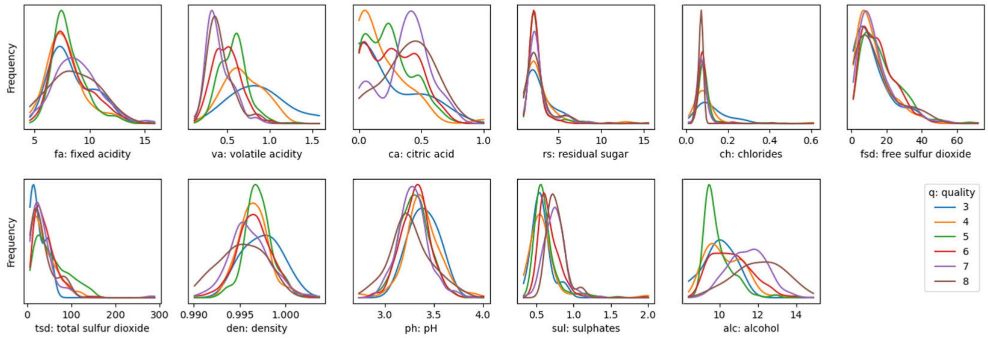

FIGURE 1. Probability density function of the values of each physicochemical property for the red wine dataset, considering different quality levels.

TABLE 1. Physicochemical features characterizing each wine sample.

Acronymous	Feature	Units[a]
Fa	Fixed acidity	g(tartaric acid)/dm^3
Va	Volatile acidity	g(acetic acid)/dm^3
Ca	Citric acid	g/dm^3
Rs	Residual sugar	g/dm^3
Ch	Chlorides	g(sodium chloride)/dm^3
fsd	Free sulfur dioxide	mg/dm^3
tsd	Total sulfur dioxide	mg/dm^3
den	Density	g/cm^3
ph	pH	-
sul	Sulphates	g(potassium sulphate)/dm^3
alc	Alcohol	vol. %

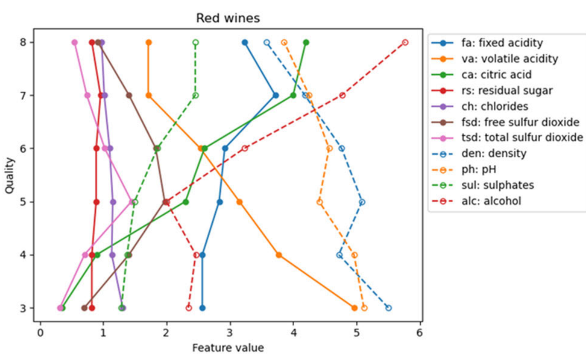

FIGURE 2. Relationship between physicochemical properties and perceived quality of red wines. The feature values are normalized in the 0-10 range and represent the median of the values with the same quality.

relationship. To overcome this difficulty, and just for graphic representation purposes, each of the quality values is depicted as a function of the median value of the features for the wines with that quality. For a better comparison, the features values are converted to the range 0 (minimum) to 10 (maximum), the same range used for the perceived quality. The result is depicted in Fig.2. This graphic provides a good perspective on the intensity and direction of the influence of each physicochemical property on the perceived quality. However, to determine the importance of each feature a single metric would be advisable.

B. THE CONCEPT OF IMPORTANCE

The importance of a feature is a concept that can be intuitively understood but needs to be elucidated more formally. Importance is a variable that can be evaluated referring to different mathematical and statistical indicators [26]. In our context, a physicochemical feature of a wine is said to be important if, by changing its value, a different perceived quality is obtained. To formalize and quantify the concept of importance, and then summarize it in a single metric, two main basic approaches have been proposed: regression and correlation.

1) IMPORTANCE BASED ON REGRESSION

A regression analysis models the perceived quality q as a function of the set of d physicochemical features $\boldsymbol{x} = (x_1, x_2, \cdots, x_d)$, according to the expression

$$q = f(\boldsymbol{x}) + \nu, \tag{1}$$

where ν is a random variable (a noise) modelling the influence of other unknown factors. Then, the influence of the j-th feature can be modelled by the function

$$r_j(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial x_j}. \tag{2}$$

This influence is different for each value of $\boldsymbol{x}$, that is, for each sample in the dataset. For the i-th sample, the influence can be formulated as

$$r_j^{(i)} = \left[\frac{\partial f(\boldsymbol{x})}{\partial x_j}\right]_{\boldsymbol{x}=\boldsymbol{x}^{(i)}} \approx \frac{\Delta q^{(i)}}{\Delta x_j^{(i)}}, \tag{3}$$

which is an estimation of the change in the quality of the i-th sample that can be expected using a small change in the value of its j-th feature. So, regression models provide local measures of the influence, that is, individual values for each sample. If a global indication of the influence of the j-th feature is required, the different values of local $r_j^{(i)}$

can be statistically summarized, for instance, using the mean value. In this case, the global influence of the j-th feature considering the n samples of the dataset can be expressed as

$$R_j = \frac{1}{n}\sum_{i=1}^{n} r_j^{(i)}. \tag{4}$$

The simplest and probably most used regression analysis is based on a multiple linear regression model; see, for instance, [34]. In this case the estimated quality of the i-th sample, $\hat{q}^{(i)}$, is modelled as the linear combination of their physicochemical features, according to the expression

$$\hat{q}^{(i)} = f(\boldsymbol{x}) = \sum_{j=1}^{d} w_j x_j^{(i)}, \tag{5}$$

where w_j is the weight of the j-th feature in the linear regression. The error of the estimation on the i-th sample is $\varepsilon^{(i)} = \hat{q}^{(i)} - q^{(i)}$, and the total loss for the dataset is defined as the mean square of these errors (MSE), obtained as

$$J = \frac{1}{n}\sum_{i=1}^{n} {\varepsilon^{(i)}}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{q}^{(i)} - q^{(i)}\right)^2. \tag{6}$$

Then, the values of the weights w_j are those that minimizes the total loss J. In this model, the influence for the i-th sample is

$$\begin{aligned} r_j^{(i)} &= \left[\frac{\partial f(\boldsymbol{x})}{\partial x_j}\right]_{\boldsymbol{x}=\boldsymbol{x}^{(i)}} = \left[\frac{\partial \sum_{j=1}^{d} w_j x_j}{\partial x_j}\right]_{\boldsymbol{x}=\boldsymbol{x}^{(i)}} \\ &= \left[w_j\right]_{\boldsymbol{x}=\boldsymbol{x}^{(i)}} = w_j, \forall i, \end{aligned} \tag{7}$$

that is, it is the same for all the samples. Then, the local and global influence have also the same value: $R_j = w_j = r_j^{(i)}, \forall i$.

If the feature values and the quality are normalized in the same range of values, then the regression coefficients are in the range $[-1, 1]$. But the feature x_j is also important to improve the quality q if its corresponding weight w_j is negative. In fact, a negative coefficient means that, by increasing the value of the feature, the quality decreases on average, meaning that it is important to consider this feature (decreasing its value) to improve the quality. Then, an important result of this subsection, is that the importance I_j of the j-th feature could be measured in the linear case using the absolute value of the regression coefficient, that is, $I_j = |w_j|$, which is in the range $[0, 1]$.

2) IMPORTANCE BASED ON CORRELATION

Using regression models to determine the importance requires a previous estimation of the function f, which is a cumbersome task except for the simplest cases (as in the linear regression). As an alternative, the correlation approach avoids the necessity to determine f. The term correlation indicates a relationship or interdependence between two variables. It is commonly employed in statistics, where several tools have been designed to grab this concept. In the case of the j-th feature x_j and the perceived quality q, the relationship can be expressed as

$$q = f_j(x_j) + \nu_j, \tag{8}$$

where f_j is an unknown function describing the influence of x_j on q, and ν_j is a random variable (a noise) modelling the influence of the remaining known features and other unknown factors. The highest correlation corresponds to a target variable q completely determined by the independent variable x_j, that is, the case when $\nu_j = 0$ and then $q = f_j(x_j)$. On the contrary, the lowest correlation occurs when the variable x_j has no influence on q, that is, when they are statistically independent: $q \perp x_j$.

Probably the most popular metric used in science and engineering for measuring the relationship between two variables is the Pearson's correlation coefficient ϱ_P [22], which is in the range $[-1, +1]$. So, in the case of the j-th feature x_j and the perceived quality q, its relationship can be measured using the expression

$$\varrho_P(x_j, q) = \frac{\operatorname{cov}(x_j, q)}{\sqrt{\operatorname{var}(x_j)\operatorname{var}(q)}}. \tag{9}$$

As for the regression, also the correlation coefficient can be negative. Then, an important result of this subsection, is that the importance I_j of the j-th feature could be measured in the linear case using its absolute value, that is, $I_j = |\varrho_P(x_j, q)|$, which is in the range $[0, 1]$.

3) IMPORTANCE FOR THE WINE DATASETS

In order to compare several metrics of the importance, it is convenient to convert them to a common range. In this paper it is proposed to use the same range that the quality's, that is, the range $[0, 10]$. So the normalized importance $\mathbb{I}_j$ of the j-th feature on the perceived quality q can be defined as

$$\mathbb{I}_j = 10\frac{I_j - \inf I_j}{\sup I_j - \inf I_j}, \tag{10}$$

where $\sup I_j$ and $\inf I_j$ are the supremum and infimum values of I_j. In the case of using the regression coefficients or the Pearson's correlation coefficients, $\sup I_j = 1$ and $\inf I_j = 0$, then $\mathbb{I}_j = 10 I_j$. Applying this metric to the red wine dataset, the importance value $\mathbb{I}_j$ of every feature are depicted in Fig.3. The values of the importance so obtained, are compatible in most cases with the results shown in Fig.1 and Fig.2. However, if the relationship between feature and quality is not very linear, the metric obtained shows a lower importance value. For instance, although citric acidity shows in those figures a similar influence on quality than volatile acidity, its importance metric is significantly lower. Even more, the free sulfur dioxide has a very low importance metric while the quality is significantly influenced by its value.

C. PERFORMANCE EVALUATION OF THE IMPORTANCE METRICS

As the Pearson's correlation coefficients and linear regression coefficients have shown important biases for non-linear

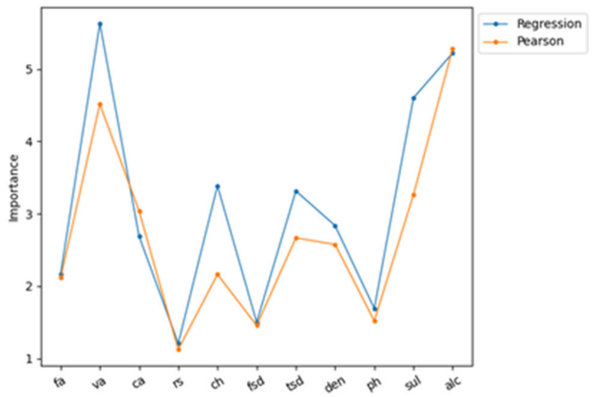

FIGURE 3. Importance of every physicochemical feature obtained using linear regression coefficients and Pearson's correlation coefficients.

relationships, it is advised to use another metrics. To explore and compare these alternatives methods, a collection of synthetic datasets is proposed using 8 types of relationships between a dependent variable q, an independent variable x_j, and a noise ν. The independent variable and noise are uniformly distributed in the range $[-1, +1]$. The dependent variable is obtained using the expression

$$q = (1-\beta) f_j\left(x_j\right) + \beta\nu. \tag{11}$$

The value of β, in the range $[0, 1]$, indicates how much noise is added. The relationships $f_j\left(x_j\right)$ used are: linear, logarithmic, cubic, quadratic, sinusoidal, piecewise, semielliptical and vertical sine. All of them are built in a way that $f_j\left(x_j\right)$ is also in the range $[-1, +1]$. Finally, the independent and dependent variables are converted to the $[0, 10]$ range, using expressions similar to Eq. 10. The resulting synthetic datasets are depicted in Fig. 4 for $\beta = 0$.

In the set previously described there are linear, monotonic non-linear (logarithmic and cubic), non-monotonic (quadratic, sinusoidal and piecewise), and multivalued (semielliptical and vertical sine) functions. The impact of adding noise to three of these examples is depicted in Fig. 5 for five values of β.

To evaluate and compare the metrics of attribute importance, three performance indicators will be employed: power, equitability and linearity. They are described in the next subsections where the Pearson-based importance metric is used as an example.

1) POWER

The first performance indicator of an importance metric derives from the concept of the statistical power of a hypothesis test [35].

It has been stated above that the importance $I_j = I\left(x_j, q\right)$ measures the dependence of the quality q on the values of the feature x_j. But also, this metric can be used in a hypothesis test to determine if x_j and q are independent or not.

In first place, let us consider a feature z that has no influence on the quality q, that is, $z \perp q$. An individual experiment takes n random samples of values $\left(z^{(i)}, q^{(i)}\right)$, yielding an importance metric $I_{ind} = I(z, q)$. Repeating the experiment several times a set of importance metrics are obtained. These values are statistically distributed according to a probability distribution function $F_z(I_z)$. For a confidence level of α, the critical value of the importance metric, I_{cr}, is defined as the value fulfilling the equation $\Pr ob\left[I_{ind} > I_{cr}\right] = 1 - F_z(I_{cr}) = \alpha$.

Now, for the feature x_j, the null-hypothesis H_0 asserts that x_j and q are independent, formally, $H_0 : x_j \perp q$. As in the previous case, the importance metric $I_j = I_j\left(x_j, q\right)$ is statistically distributed according to a probability distribution function $F_j\left(I_j\right)$. Then, the power of I_j is defined as the probability of rejecting H_0, that is,

$$W_{I_j} = \text{Prob}\left[I_j > I_{cr}\right] = 1 - F_j\left(I_{cr}\right). \tag{12}$$

If q is obtained as a certain function of x_j, combined with some proportion of noise β (according (11)), then the power of I_j, W_{I_j}, can be written as a function of β, that is, as $W_{I_j}(\beta)$. The mean value of the power W_{I_j} in the interval $\beta \in [0, 1]$ can be used as an indicator of the performance of the importance metric. It is defined as

$$\bar{W}_{I_j} = \int_0^1 W_{I_j}(\beta)\, d\beta. \tag{13}$$

Additionally, it is defined the critical amount of noise attainable by an importance metric, or critical noise in short, as the proportion of noise, β_{cr}, such that the null hypothesis (x_j and q are independent) is rejected with a confidence level of α. More formally,

$$W_{I_j}(\beta_{cr}) = \text{Prob}\left[I_j > I_{cr}\right] > 1 - \alpha. \tag{14}$$

The critical noise β_{cr} is an indicator of the performance of the importance metric. It is in the range $[0, 1]$ and the greater the critical noise, the better the importance metric.

These concepts are graphically explained in Fig 6 using an example, where the Pearson-based importance metric defined in (9), and a monotonic (cubic) relationship are used. The upper left plot represents the case of a feature z that has no influence on the quality q, that is, $z \perp q$. This is equivalent to say that they do have a cubic relationship but with a 100% of noise ($\beta = 1$ in (11)). In the plot, the blue line represents the probability density function of the importance metric, and the red-shadowed area is the probability that $\Pr ob\left[I_z > I_{cr}\right] = \alpha = 5\%$. The resulting critical value is $I_{cr} = 0.064$. The two other plots on the left represents the case of a feature x_j that has cubic influence on the quality q, according to (11), with a noise level of $\beta = 0.9$ and $\beta = 0.8$ respectively. The shadowed areas represent the respective powers. Finally, the right plot depicts the power of the importance metric for different experiments with a continuous set of values of noise, $W_I(\beta)$, and a mean power $\bar{W}_I = 0.916$. The three black dots correspond to each graph on the left-hand side of the figure. The blue dot indicates the critical noise β_{cr} which results to be $\beta_{cr} = 0.839$.

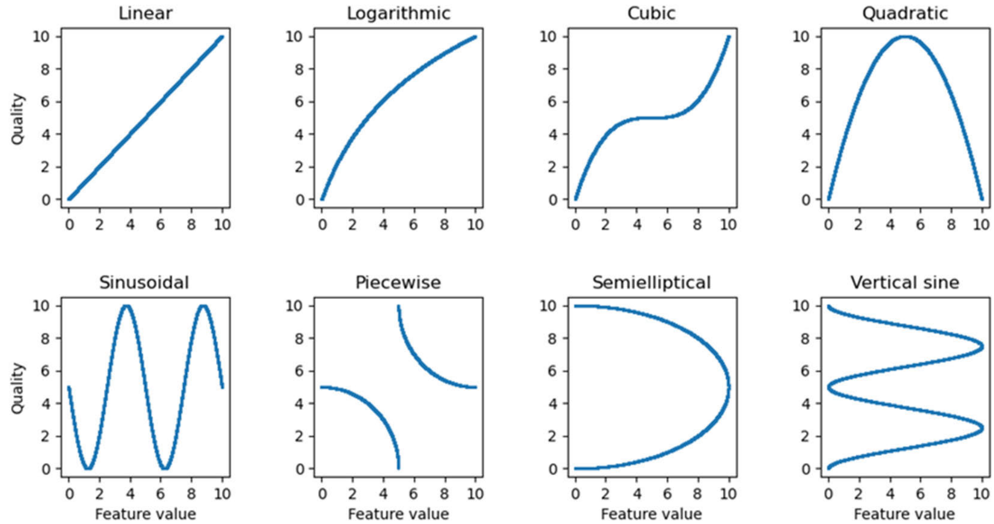

FIGURE 4. Synthetic datasets without noise ($\beta = 0$).

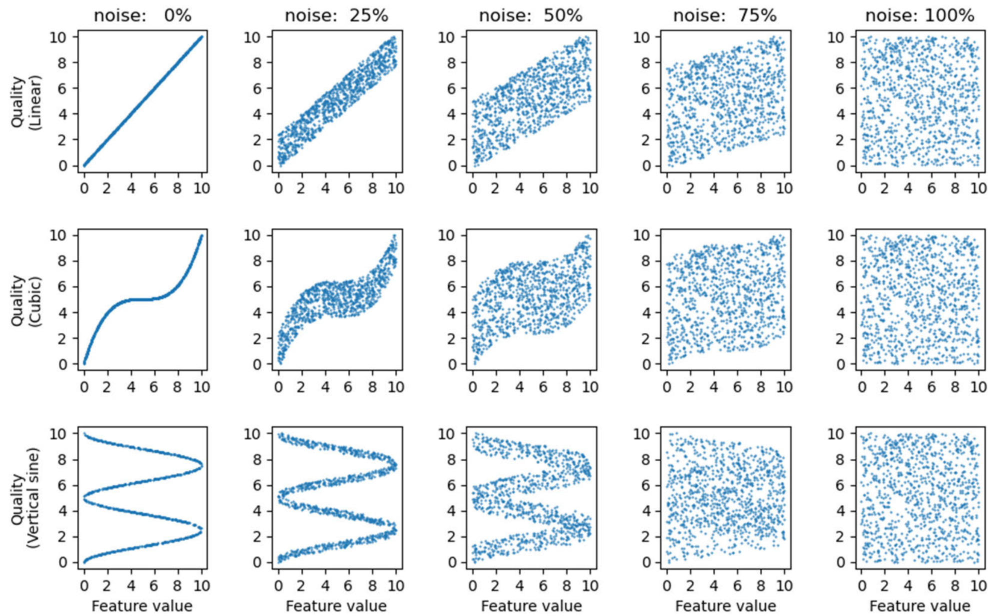

FIGURE 5. Examples of synthetic datasets for an increasing level of noise.

2) EQUITABILITY

It is said that an importance metric is fully equitable if it has the same value for different types of relationships between a dependent variable q, an independent variable x_j, with the same level of noise β [36]. Let us consider a set of relationships defined by the functions $f_k \in \mathcal{F} : q = (1 - \beta) f_k \left(x_j\right) + \beta \nu$. The importance of the feature x_j on obtaining q using the k-th relationship f_k is assessed by the importance metric $I_k \left(\beta; f_k\right)$. Then, the equitability of this metric for a certain level of noise can be defined as a function of the maximum range of its values, that is,

$$Q(\beta) = 1 - \left[\max_k I_k \left(\beta; f_k\right) - \min_k I_k \left(\beta; f_k\right)\right], \tag{15}$$

which is in the range $[0, 1]$. An example of equitability is shown in Fig. 7, where the Pearson-based importance metric defined in Eq. 9 is used. Each line depicts the evolution of the importance metric as the level of noise increases, for the 8 functions described in Fig.4. The range of values for every β is shown as a blue shadowed area where, the wider the area, the lesser the equitability $Q(\beta)$. The equitability coefficient of the importance metric (for any level of noise), κ, is defined as the minimum value of $Q(\beta)$, that is,

$$\kappa = \min_\beta Q(\beta). \tag{16}$$

If the set of relationships, $\mathcal{F}$, contains discrete multivalued functions, it is previously separated in subsets with the same

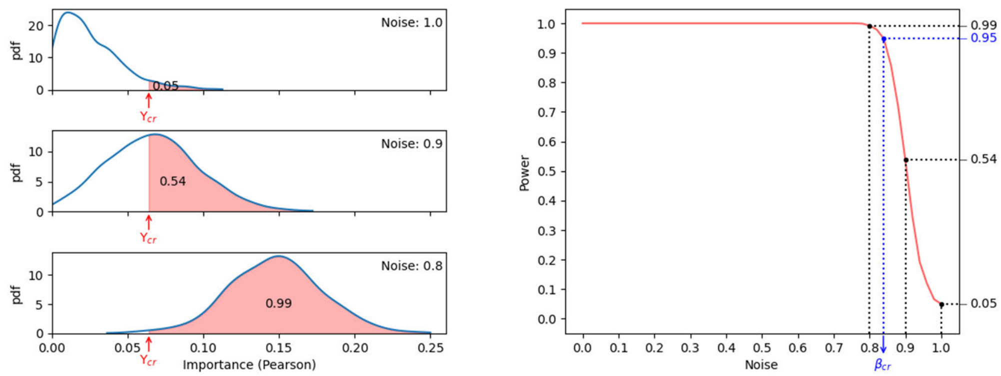

FIGURE 6. Power of the (Pearson-based) importance metric for the cubic synthetic dataset. Left: noise of 100% (top), 90% (middle), and 80% (bottom). Right: power for a continuous set of noise values.

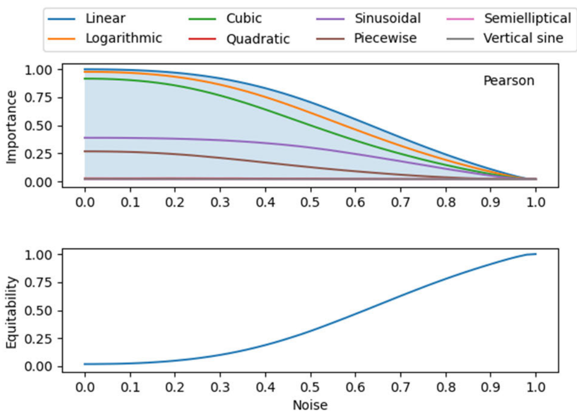

FIGURE 7. Equitability of the (Pearson-based) importance metric. Top: importance metric versus noise for the eight functions defined in section II-C. Bottom: equitability as a function of noise.

number of multiple values. For the subset containing functions with m multiples values, its equitability coefficient is computed as κ_m. Finally, the equitability coefficient of the correlation coefficient is derived as $\kappa = \min_m \kappa_m$. In the example, the equitability coefficient results to be $\kappa = 0$.

3) LINEARITY

Besides power and equitability, proposed as performance indicators in previous works, we also consider that a perfect metric of importance should have three additional characteristics:

1) Be linear with respect to the noise. That is, it is desirable that the relationship between the amount of randomness in the dataset (the noise coefficient β in Eq. 11) and the importance metric $I_j(\beta)$, be as much linear as possible.
2) Have the minimum value for completely noisy datasets, that is, $I_j(1) = 0$.
3) Have the maximum value for noiseless datasets. This maximum value is discussed below.

The highest importance corresponds to a target variable q completely determined by the independent variable x_j, that is, the noiseless case when $\nu_j = 0$ and then $q = f_j(x_j)$. If the concept of importance is mathematically defined by a function in the range [0, 1] (as in Pearson's, and for most of the correlation coefficients), then the straightforward answer for the quest about its maximum value is $I_j = 1$. But there are certain particular cases where the intuitive concept of importance should not have the maximum value in the noiseless case. Let us examine them in detail.

Consider a target variable q, which can take only one out of L_q discrete levels (in the wine datasets $L_q = 11$), and a relationship with the u-th feature $f_u : q = f_u(x_u)$, which is a single valued function, that is, only one value of q corresponds to each value of x_u. In this noiseless scenario, x_u fully determines the quality q. Then, the importance of this feature should be $I_u = 1$.

Consider now the v-th feature whose noiseless relationship with the quality is defined by $f_v : q = f_v(x_v)$, which is a multivalued function with m_j values of q for each value of x_v. The free sulfur dioxide in the red wine dataset is an example of such multivalued function ($m_j = 2$) as it is shown in Fig. 2. As a value of x_v determines m_v out of L_q possible values, then x_v is not so determinant to define the quality q, as x_u is. Then, a good metric of its noiseless importance should be $I_v < I_u = 1$. This effect can be expressed adding a correcting term, θ_v, to the importance such that $I_v = 1 - \theta_v$.

The correction term should be $\theta_v = 0$ for the single valued functions. And in the other extreme, for a feature x_w defined by a multivalued function f_w with $m_w = L_q$, choosing a value for x_w determines L_q possible values of q, which is equivalent to a random selection of the quality. So, the importance of this feature should be $I_w = 0$, which is equivalent to say that the corresponding correction term should be $\theta_w = 1$. The simplest function which fulfills those conditions is a piecewise linear function defined as

$$\theta_j = \max\left(0, \frac{m_j - 1}{L_q - 1}\right). \tag{17}$$

To clarify the concept of linearity, two examples of importance metrics are applied to the linear and to the semielliptical synthetic datasets with different levels of noise. The result is depicted as a blue line in Fig. 8 where it is compared with an optimal metric $I_j^*(\beta)$ (dashed line). The semielliptical is

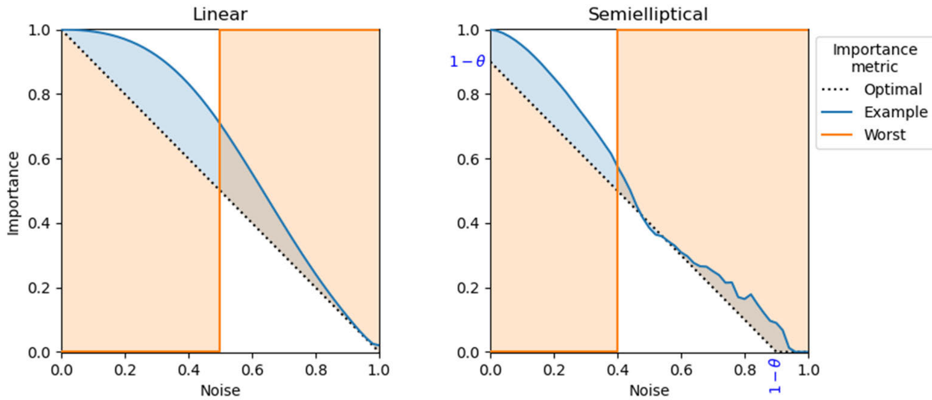

FIGURE 8. Linearity of the importance metric. An example importance metric (blue line) is compared to the optimal (dashed line) and worst (orange line) possible metrics. The areas between each metric and the optimal one, have been filled. Left: single valued linear function; Right: multivalued semielliptical function. Power of the (Pearson-based) importance metric for the cubic synthetic dataset.

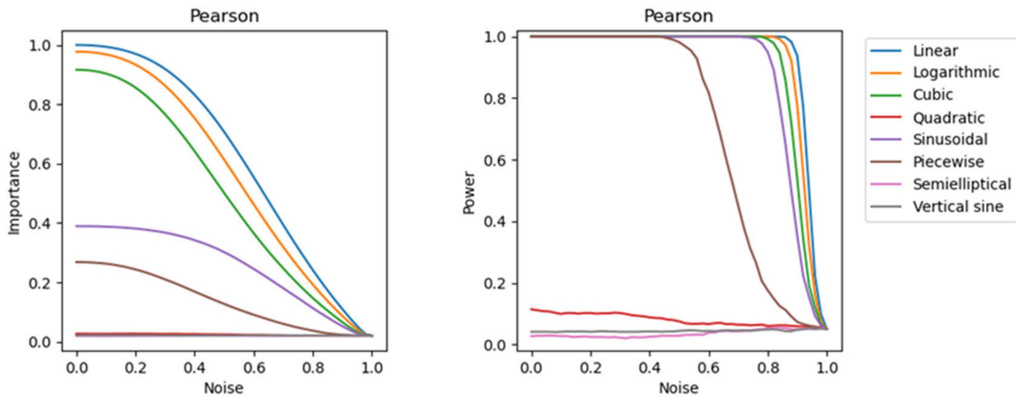

FIGURE 9. Left: Pearson's-based importance for the synthetic datasets as a function of the noise level (β). Right: Power of the importance metric.

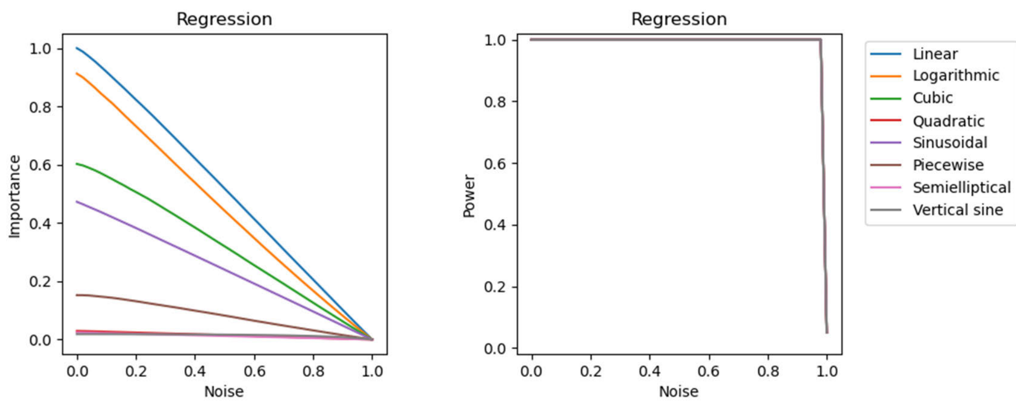

FIGURE 10. Left: Importance based on the regression coefficient for the synthetic datasets as a function of the noise level (β). Right: Power of the importance metric.

a discrete multivalued function with $m_j = 2$ and $L_q = 11$. Then, for this example, the optimal noiseless importance is

$$I_j^*(0) = 1 - \frac{m_j - 1}{L_q - 1} = 1 - \frac{2-1}{11-1} = 0.9. \quad (18)$$

To assess the linearity of the example metric, the difference between $I_j(\beta)$ and $I_j^*(\beta)$ is considered. In particular, the area between these two functions (the blue region in Fig. 8) provides a good understanding of this concept. This area is defined by

$$A_j = \int_0^1 \left| I_j(\beta) - I_j^*(\beta) \right| d\beta. \quad (19)$$

It has an opposite meaning that performance, as the lower the value of A_j, the higher the performance. Also, it can be easily derived that the maximum value of A_j, corresponding to the worst case of the importance metric (the orange line and region in Fig. 8, is

$$\max A_j = \frac{3-2\theta^2}{4}. \quad (20)$$

Then, to obtain an indicator in the range [0, 1], a linearity coefficient is defined as

$$\lambda_j = 1 - \frac{A_j}{\max A_j} = 1 - \frac{4}{3-2\theta^2}\int_0^1 \left| I_j(\beta) - I_j^*(\beta) \right| d\beta. \quad (21)$$

In the examples, the linearity coefficient results to be $\lambda_j = 0.835$ (left) and $\lambda_j = 0.897$ (right).

D. IMPORTANCE METRICS FOR LINEAR RELATIONSHIP

By applying the Pearson's-based importance to the synthetic datasets, the results depicted in Fig. 9 are obtained. It can be clearly seen that this definition of importance shows three problems. Firstly, for the linear dataset, importance does not linearly decrease as the noise increases. For instance, for 25% noise ($\beta = 0.25$) an importance $I_j = 0.95$ is obtained (only a 5% decrease in the importance metric). Secondly, the noise-free ($\beta = 0$) non-linear monotonic (logarithmic and cubic) relationships do not obtain the maximum importance, $I_j(0) \neq 1$. And finally, the non-monotonic functions show a low (or very low) importance for any value of the noise level, β. Alternatively, if the power of the importance metric is considered, Pearson's based importance also shows a low or very low performance for several relationships in the synthetic dataset.

Similar results are obtained when the regression coefficient is used to define the importance metric (Fig.10). In comparison to Pearson's based metrics, they show a more linear behavior with respect to the noise and a high power, but significantly lower values for non-linear datasets.

III. IMPROVED IMPORTANCE METRICS

A. IMPORTANCE METRICS FOR NON LINEAR RELATIONSHIP

The previous sections have clearly demonstrated the low performance of the Pearson's and regression-based importance I_j, when they are applied to the synthetic datasets. The problems of Pearson's correlation and regression coefficients to properly identify correlations in non-linear functions are widely known and have been extensively documented [23]. To tackle these problems several solutions have been proposed. Among them, Spearman's correlation coefficient [37] is probably one of the first and best-known methods. It relies on obtaining the Pearson's correlation coefficient between the rank values of the variables, instead of their values. The rank value of the i-th value of a variable, $\text{rank}\left(z^{(i)}\right)$, is defined as its index after ordering the complete set of values of that variable of z. As the Pearson's, also Spearman's is in the range $[-1, 1]$. So, the importance based on the Spearman's correlation coefficient between the j-th feature x_j and the quality q, can be defined as

$$I_j = \left|\varrho_S(x_j, q)\right| = \left| \frac{\text{cov}\left[\text{rank}(x_j), \text{rank}(q)\right]}{\sqrt{\text{var}\left[\text{rank}(x_j)\right]\text{var}\left[\text{rank}(q)\right]}} \right|. \quad (22)$$

The results obtained for the synthetic datasets are shown in Fig. 11. It can be seen that, now, linear and non-linear monotonic (logarithmic and cubic) relationships do obtain the maximum importance, $I_j(0) = 1$, for noise-free datasets. However, the relationships in the remaining datasets are not properly reflected by Spearman's metric.

Even for monotonic relationships, Spearman-based importance does not linearly decrease as the noise increases. To obtain a better linearity in those cases, the Kendall's correlation coefficient [38] has been proposed. The u-th pair of values $\left(x_j^{(u)}, q^{(u)}\right)$ and the v-th pair of values $\left(x_j^{(v)}, q^{(v)}\right)$, where $u < v$, are said to be concordant if $\left[\left(x_j^{(u)} < x_j^{(v)}\right) \wedge \left(q^{(u)} < q^{(v)}\right)\right] \vee \left[\left(x_j^{(u)} > x_j^{(v)}\right) \wedge \left(q^{(u)} > q^{(v)}\right)\right]$. Calling n_c and n_d the number of concordant and discordant pairs respectively, the Kendall's correlation coefficient is defined as

$$\tau(x_j, q) = \frac{n_c - n_d}{n_c + n_d}. \quad (23)$$

To solve the tied pairs, that is, those where either $x_j^{(u)} = x_j^{(v)}$ or $q^{(u)} = q^{(v)}$, several variants have been proposed. In this research the τ_b formulation is used [39]. As the Kendall's correlation coefficient is in the range $[-1, 1]$, the importance based on it is defined as $I_j = \left|\tau_b(x_j, q)\right|$.

The results obtained for the synthetic datasets are shown in Fig.12. They are like those obtained using Spearman's, but the linearity for monotonic relationships has significantly been improved.

The Spearman's and Kendall's correlation coefficient extend the usability of Pearson's to non-linear monotonic relationships by using some rank-based functions of the values $\left(x_j^{(i)}, q^{(i)}\right)$ instead of the values themselves [40]. The maximal correlation coefficient extends this idea by exploring all real-valued functions with zero mean and finite variance, $f_{x_j}(x_j)$ and $f_q(q)$, and applying the Pearson's correlation coefficient to the new transformed variables. The maximal correlation coefficient is then the supremum of the set of Pearson's correlation coefficients for every f_{x_j} and f_q [41]. More formally, it is defined as

$$\varrho_M(x_j, q) = \sup_{f_{x_j}, f_q} \varrho_P\left[f_{x_j}(x_j), f_q(q)\right]. \quad (24)$$

The direct calculation of this value is usually very difficult [42]. Instead, a good estimation can be obtained using the Alternating Conditional Expectations (ACE) algorithm [43].

The results obtained for the synthetic datasets are shown in Fig.13. It can be seen that now, even non-monotonic

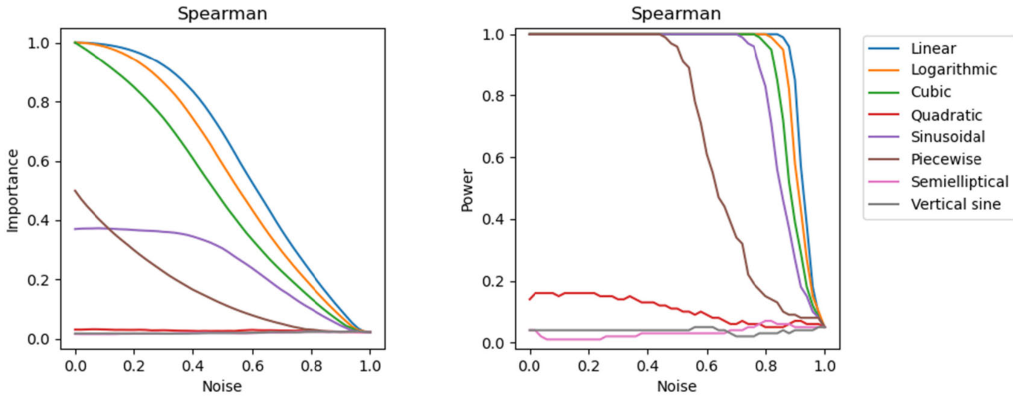

FIGURE 11. Left: Importance based on the Spearman's correlation coefficient for the synthetic datasets as a function of the noise level (β). Right: Power of the importance metric.

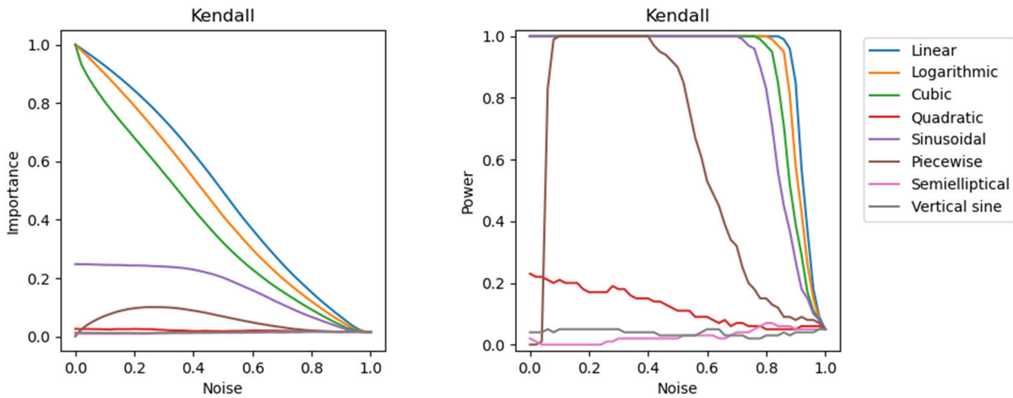

FIGURE 12. Left: Importance based on the Kendall's correlation coefficient for the synthetic datasets as a function of the noise level (β). Right: Power of the importance metric.

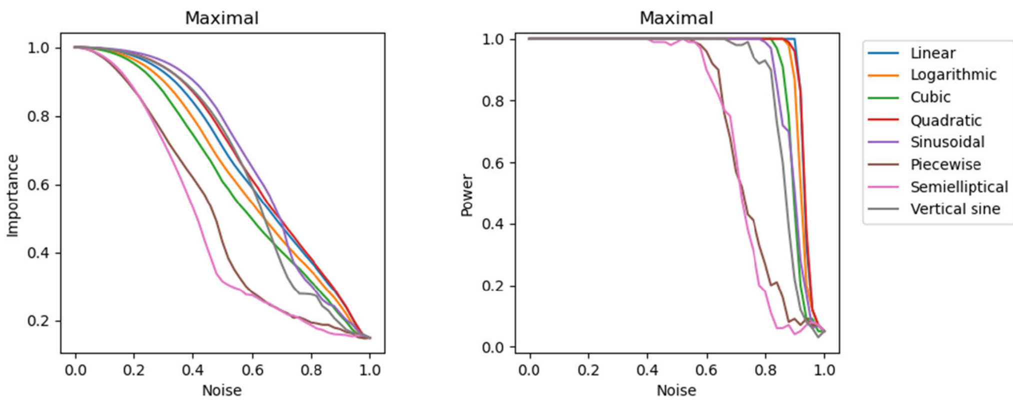

FIGURE 13. Left: Importance based on the maximal correlation coefficient for the synthetic datasets as a function of the noise level (β). Right: Power of the importance metric.

relationships, are properly captured by the metric. For the case of noise-free relationships, maximal correlation obtains a convenient value $\varrho_M\left(x_j, q\right) = 1$. However, for q completely independent of x_j (which is equivalent to $\beta = 1$), maximal correlation is not null, but $\varrho_M\left(x_j, q\right) = \varrho_{M_{min}} \neq 0$. This drawback can easily be overcome by normalizing its value. Then, the importance metric can be defined as

$$I_j = \frac{\varrho_M\left(x_j, q\right) - \varrho_{M_{min}}}{1 - \varrho_{M_{min}}}, \tag{25}$$

which is in the range [0, 1].

Although this metric properly behaves for any type of relationships, it lacks some linearity as defined in Section III).

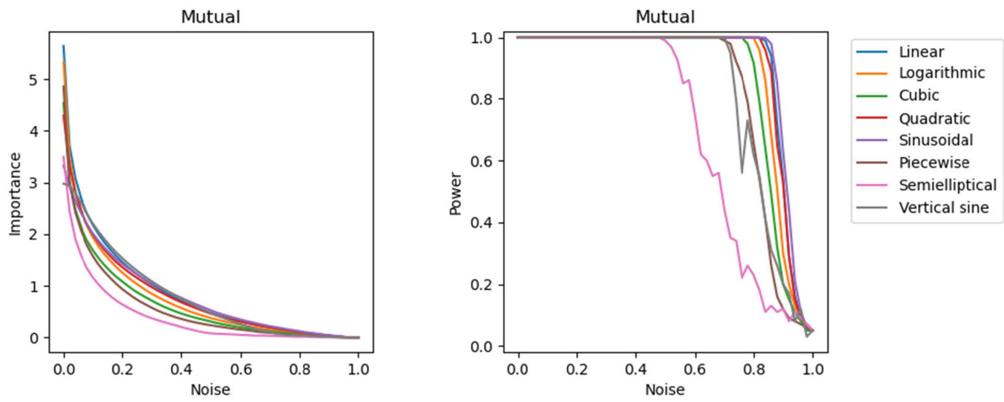

FIGURE 14. Left: Importance based on the mutual information for the synthetic datasets as a function of the noise level (β). Right: Power of the importance metric.

Also, the ACE algorithm requires a very high computational effort in comparison with other alternatives. This problem will be addressed in more detail in Section 0.

B. IMPORTANCE METRICS BASED ON INFORMATION THEORY

In previous sections, the relationship between a certain feature and the perceived quality has been addressed using either linear regression or some kind of statistical correlation. An alternative to that approach is to look for the metric of the relationship in the information theory [44]. A first metric based on this methodology is the mutual information [45] of the variables x_j and q, defined as

$$MI_j = MI\left(x_j, q\right) = \int_{x_j \in \mathcal{X}_j} \int_{q \in Q} p\left(x_j, q\right) \log \frac{p\left(x_j, q\right)}{p\left(x_j\right) p\left(q\right)} dx_j dq. \quad (26)$$

The results obtained for the synthetic datasets are shown in Fig. 14. It can be seen that all relationships, even non-monotonic ones, show a similar behavior. For the case of q completely independent of x_j ($\beta = 1$) mutual information obtains a convenient value $MI_j = 0$. However, for noise-free relationships ($\beta = 0$) maximal correlation is greater than one, $MI_j > 1$.

The straighforward way to compute the mutual information of two variables, x_j and q, relies on, firstly, estimate the probabilities $p\left(x_j\right)$, $p\left(q\right)$ and $p\left(x_j, q\right)$. And this estimation can be obtained by dividing the $\left(x_j, q\right)$ plane in a grid of size $N_x \times N_q$. Let us count the samples in each bin, calling n_u the number of them in the u-th row, n_v their number in the v-th column, and n_{uv} the samples in the uv-th cell. For a total number of samples n, the mutual information can be approximated by its binned version defined as

$$MI\left(x_j, q\right) \approx MI_b\left(x_j, q\right) = \sum_{u=1}^{N_x} \sum_{v=1}^{N_q} MI_{uv}, \quad (27)$$

where

$$MI_{uv} = \sum_{u=1}^{N_x} \sum_{v=1}^{N_q} p_{uv} \log \frac{p_{uv}}{p_u p_v} \approx \sum_{u=1}^{N_x} \sum_{v=1}^{N_q} \frac{n_{uv}}{n} \log \frac{\frac{n_{uv}}{n}}{\frac{n_u}{n} \cdot \frac{n_v}{n}}. \quad (28)$$

If mutual information is obtained using this approach, then it can be converted to the range $[0, 1]$ using the Maximal Information Coefficient (MIC) [46] defined as

$$MIC\left(x_j, q\right) = \max_{N_x \times N_q} \frac{\max_{u,v} MI_{uv}}{\log\left(\min\left\{N_x, N_q\right\}\right)}. \quad (29)$$

However, bin-based estimations of mutual information have been proved to show systematics errors caused, first, by approximating MI by its binned version MI_b and, second, by approximating the probabilities by the frequency ratios [47].

Alternatively, this research follows a different approach based on the k-nearest neighbours (kNN) which clearly outperforms bin-based approaches [48]. Then, instead of using the MIC, mutual information is normalized by dividing its value by the maximum MI, which is obtained as the mutual information of a feature with itself, that is, $MI\left(x_j, x_j\right)$. Then, an importance metric can be derived by the expression

$$I_j = \frac{MI\left(x_j, q\right)}{MI\left(x_j, x_j\right)}, \quad (30)$$

which is in the range $[0, 1]$. The normalizing divisor, $MI\left(x_j, x_j\right)$, can be obtained as the $MI\left(x_j, q\right)$ when q has a noise-free ($\beta = 0$) linear relationship with x_j, that is, $q = x_j$. This value corresponds to the blue line for $\beta = 0$ in Fig.14, $MI\left(x_j, x_j\right) = 5.651$ for the synthetic datasets described in Section 0.

Although the range problem has been fixed by normalizing the mutual information, it still shows a significant non-linear behavior. So, another way of normalizing its value, while seeking a more linear dependence of noise (β), mutual information can be transformed into the Linfoot's information

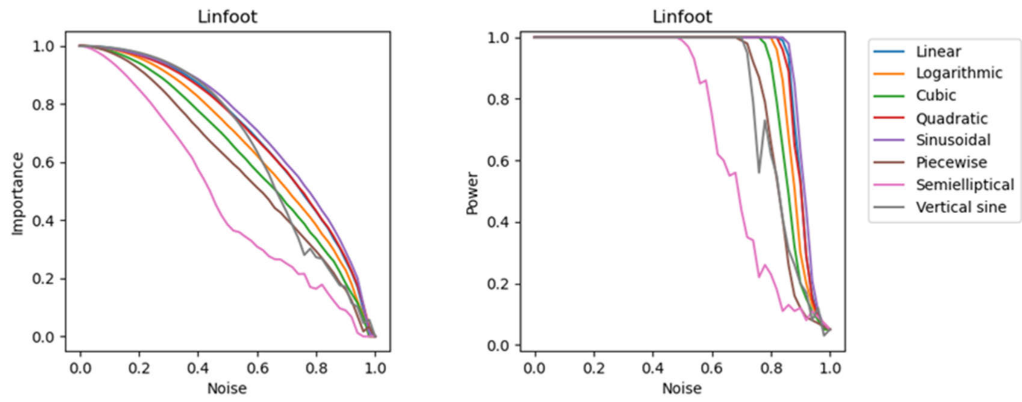

FIGURE 15. Left: Importance based on the Linfoot's information coefficient of correlation for the synthetic datasets as a function of the noise level (β). Right: Power of the importance metric.

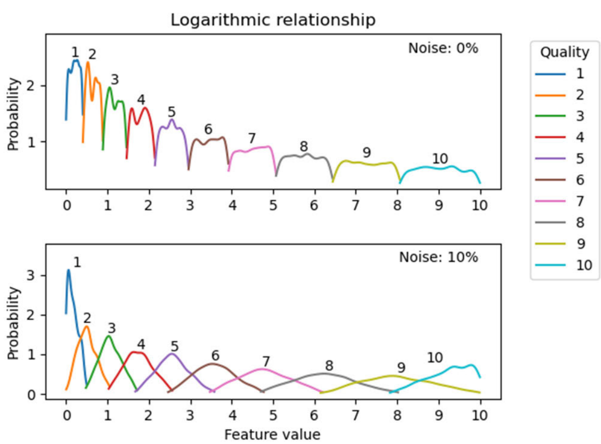

FIGURE 16. Probability density functions of the feature values for the 10 possible values of the target variable (the perceived quality q), using the synthetic dataset with a noisy logarithmic relationship. Top: noise-free dataset ($\beta = 0$). Bottom: noise level, $\beta = 10\%$.

coefficient of correlation [49], defined as

$$\varrho_L\left(x_j, q\right) = \sqrt{1 - e^{-2 \cdot MI\left(x_j, q\right)}}, \tag{31}$$

which is in the range [0, 1]. The results obtained applying the corresponding importance metric, $I_j = \varrho_L\left(x_j, q\right)$, to the synthetic datasets are shown in Fig.15. It can be seen that: all relationships are properly modelled (even non-monotonic ones); in the case of q completely independent of x_j ($\beta = 1$), it is obtained a convenient value $I_j = 0$; and for noise-free relationships ($\beta = 0$), $I_j = 0$. Also, the linearity of the dependence of noise is significantly better than for the mutual information.

C. THE JENSEN-SHANNON DIVERGENCE

None of the previously described methods have successfully met the power, equitability, and linearity criteria for importance metrics. For this reason, in this research an alternative importance metric is proposed based on the consideration that quality q has discrete values. Let us consider the same synthetic datasets described in Section 0 but now assuming that q is an integer in the range [0, 10]. As an example, the probability density functions corresponding to the values of x_j for each value of q, are depicted in Fig.16. They have been obtained for a synthetic dataset with 10,000 samples and a logarithmic relationship with no noise (top), and a level of noise $\beta = 0.1$ (bottom). In the noise-free example the pdfs corresponding to each value of q are clearly separated. On the other hand, the noisy example shows a certain overlap between pdfs. Then, the level of overlapping may be used as an indication of the independence between variables: the larger is the divergence between pdfs, the more intense the relationship.

To formalize the divergence between two pdfs, the Jensen-Shannon Divergence (JSD) is employed [24]. Calling p_u to the pdf of the samples $x_j^{(i)}$ such that $q^{(i)} = u$ and, analogously, p_v to the pdf of the samples $x_j^{(i)}$ such that $q^{(i)} = v$, the JSD of these two pdfs is defined as

$$JSD\left(p_u, p_v\right) = \frac{1}{2} KLD\left(p_u, p_w\right) + \frac{1}{2} KLD\left(p_u, p_w\right); p_w = \frac{p_u + p_v}{2}, \tag{32}$$

where KLD represents the Kullback-Leibler Divergence [50] obtained as

$$KLD\left(p_u, p_w\right) = \int_{-\infty}^{\infty} p_u\left(x_j\right) \log_2\left[\frac{p_u\left(x_j\right)}{p_w\left(x_j\right)}\right] dx_j. \tag{33}$$

It should be noted that the JSD is always in the range [0, 1]. Based on these expressions, the importance metric is defined as the mean of the JSD for every pair of pdfs,

$$\varrho_J\left(x_j, q\right) = \frac{2}{n_q\left(n_q - 1\right)} \sum_{u=1}^{n_q - 1} \sum_{v=u+1}^{n_q} JSD\left(p_u, p_v\right). \tag{34}$$

Applying these definitions to the synthetic datasets, the results depicted in Fig.17 are obtained. It can be observed that JSD shows a high power for all the relationships between feature and quality. It also has a very high linearity for most

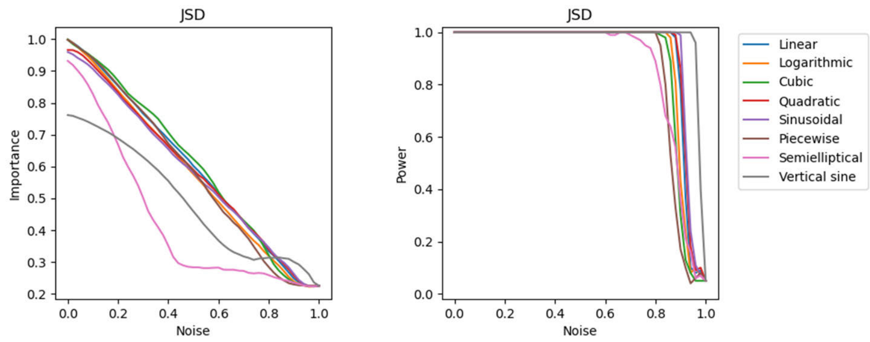

FIGURE 17. Left: Importance based on the Jensen-Shannon Divergence (JSD) for the synthetic datasets as a function of the noise level (β). Right: Power of the importance metric.

functions, with a not so good performance in the cases of multivalued functions (semielliptical and vertical sine) which will be later addressed. Finally, JSD has a value $\varrho_J \approx 1$ for most of the noiseless cases ($\beta = 0$), but a value $\varrho_J = \varrho_{J_{min}} > 0$ for the case of no relationship between quality and feature values ($\beta = 1$). Then, it is easy to obtain an importance metric in the range [0, 1] using the expression

$$I_j = \frac{\varrho_J - \varrho_{J_{min}}}{1 - \varrho_{J_{min}}}. \tag{35}$$

In the previous analysis, it has been shown that JSD importance metric has a very good behavior. Then, this will be the importance metric employed in the next Sections, where it will be applied to the wine datasets. A more detailed comparison of importance metrics will be undertaken in Section 0.

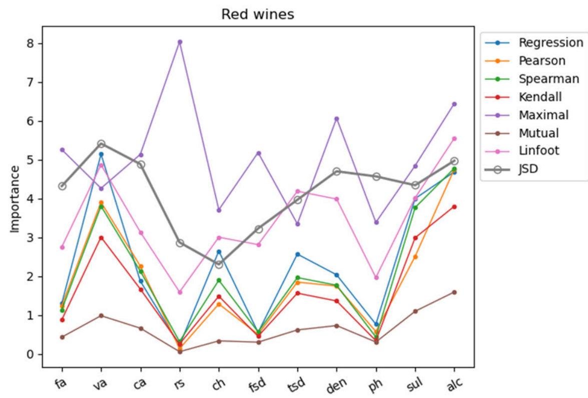

FIGURE 18. Importance metrics of the red wine dataset.

IV. RESULTS

A. THE IMPORTANT OF THE WINE PROPERTIES

Each wine feature can be characterized by its importance metrics obtained by the eight methods previously described. Then, wine features can be represented by an 11×8 matrix design X_f where each of the 11 rows defines a wine feature using a vector with the 8 importance metrics, one in each column.

The application of these eight importance metrics to the red wine dataset offers the results depicted in Fig.18. It can be seen that the JSD-based metric has a smoother evolution than the first four metrics (regression, Pearson, Spearman and Kendall). This is probably due to the fact that these metrics do not properly reflect non-monotonic and/or multivalued functions as, for instance, free sulfur dioxide (*fsd*) or *ph*. Also, the JSD-based metric has similar values for citric acidity (*ca*) and volatile acidity (*va*) fixing the problem indicated in Section III).

Fig 18 also shows that JSD and Linfoot importance metric, two of the most successful metrics in the synthetic datasets, obtain similar results for most of the features. The main exception to this rule is the *ph*, that, in Linfoot metric, has a significant lower value than density (*den*), while in JSD metric both have similar results. By comparing the dependance of quality on these two features shown in Fig 1 and Fig 2, it is clear that their importance should be similar, that is, that JSD metric more properly reflects the first sight intuition.

It has been demonstrated in the synthetic datasets that regression coefficient perfectly capture the importance of linear relationships (Fig 10), while JSD metric is able to reflect non-linear ones (Fig 17). Then, by comparing the importance metric based on JSD and regression, an indication of the linearity of the relationship between each feature and the perceived quality can be obtained. The result is shown in Fig 19 where, for instance citric acidity (*ca*) is more distant to the diagonal (that is, JSD much greater than regression coefficient) than alcohol (*alc*). Then a more non-linear behavior should be expected on *ca* than on *alc*, which is shown in Fig. 2. In this plot it is clear than the most important features are *va*, *alc* and *sul*.

By reducing the dimensionality of the 11×8 matrix design X_f to 11×2, each wine feature can be plotted in a bidimensional (2D) graph. The result obtained using Principal Component Analysis (PCA) [51] is depicted in Fig. 20, showing a circle for each feature (each row in X_f). The information conveyed by this graph can be improved by adding biplots vectors [52], displayed as an arrow for each importance metric (each column in X_f). Th resulting biplots show that vertical component of PCA representation is mainly determined by the maximal correlation, while the remaining importance

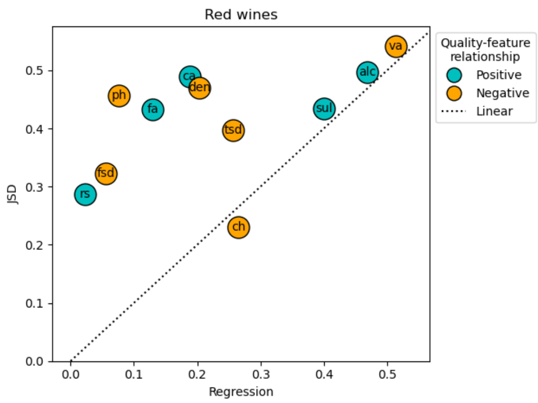

FIGURE 19. **Comparison of the importance metric for each feature of the red wine dataset. The importance is obtained using regression coefficients (horizontal) and JSD (vertical). Any noiseless linear relationship between features and the perceived quality should be in the diagonal (dotted line). The features negatively related to quality are depicted in orange.**

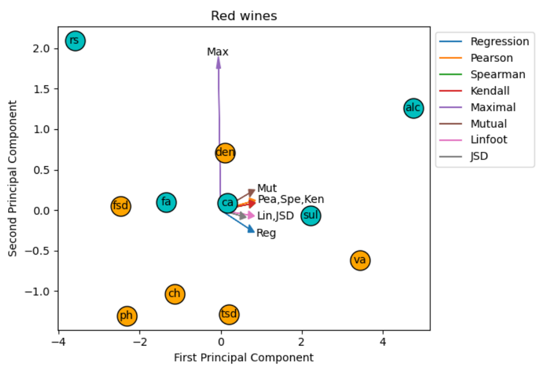

FIGURE 20. **Principal component analysis of the red wine features. Biplot vectors represent the contribution of each importance metric to its principal components. The features negatively related to quality are depicted in orange.**

metrics are approximately aligned in the horizontal direction. Then, it can be interpreted that features placed in the right-hand side of the plot (*alc*, *va* and *sul*) have an impact on quality over the mean in at least 7 importance metrics. Then, they clearly should be considered as important features.

B. IMPORTANCE PERFORMANCE ANALYSIS

The importance of a feature is the main element to be considered in a prescriptive analysis prior to make the decision about what to do to obtain an improved perceived quality. However, if the value of the feature has already been optimized, then it does not make sense to dedicate more efforts to continue improving that feature. The decision about what to do is then bidimensional.

This intuitive idea can be formalized using the importance-performance analysis (IPA) [26] where the importance is complemented with a performance dimension. The performance of the j-th feature of the i-th wine sample, if it is positively related to the quality, that is $\varrho\left(x_j, q\right) > 0$, is defined as its value $x_j^{(i)}$. To compare the performance of

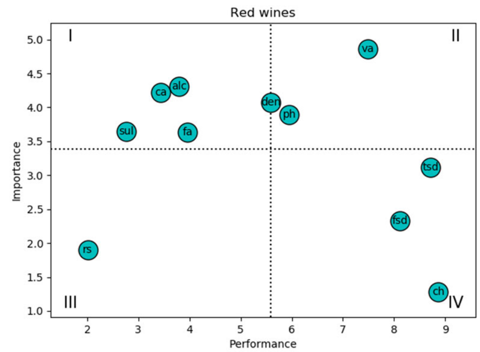

FIGURE 21. **Importance performance analysis for the red wine dataset.**

different attributes, they must be normalized to the same range, for instance to the [0, 10] range. Then, the normalized performance $P_j^{(i)}$ can be obtained by using the expression

$$P_j^{(i)} = 10 \frac{x_j^{(i)} - \min_i x_j^{(i)}}{\max_i x_j^{(i)} - \min_i x_j^{(i)}}, \forall j | \varrho\left(x_j, q\right) > 0. \quad (36)$$

But the performance of a feature is not only related to its value $x_j^{(i)}$. Indeed, for a feature negatively related to the quality, the meaning of well-performing is to have a low value of the feature. For these features the definition of performance must be inverted, obtaining the expression

$$P_j^{(i)} = 10 \frac{\max_i x_j^{(i)} - x_j^{(i)}}{\max_i x_j^{(i)} - \min_i x_j^{(i)}}, \forall j | \varrho\left(x_j, q\right) < 0. \quad (37)$$

The Pearson's correlation coefficient is used in this research to know if a feature is positively or negatively related to the quality, that is, in the previous equations $\varrho\left(x_j, q\right) = \varrho_P\left(x_j, q\right)$.

The value $P_j^{(i)}$ is called a local performance as it refers to a single wine in the dataset. The performance of a feature in the whole dataset (global performance) is defined as the mean performance of their individual elements, that is,

$$\mathbb{P}_j = \frac{1}{n} \sum_{i=1}^{n} P_j^{(i)}. \quad (38)$$

The global performance of symmetrically distributed features is just in the middle of the range, $\mathbb{P}_j = 5$ in this case. The results of applying these definitions of global performance and a JSD-based importance to the red wine datasets is depicted in Fig. 21. The mean values of importance and performance are plotted as dotted lines, dividing the plane in four quadrants. The meaning of each quadrant is detailed in Table 2.

In quadrant I, there are the wine features with a high importance (an impact on perceived quality over the mean value of importance), and a low performance (a performance under the mean value of performance). So, the features in this quadrant are the primary goal on which to focus design efforts.

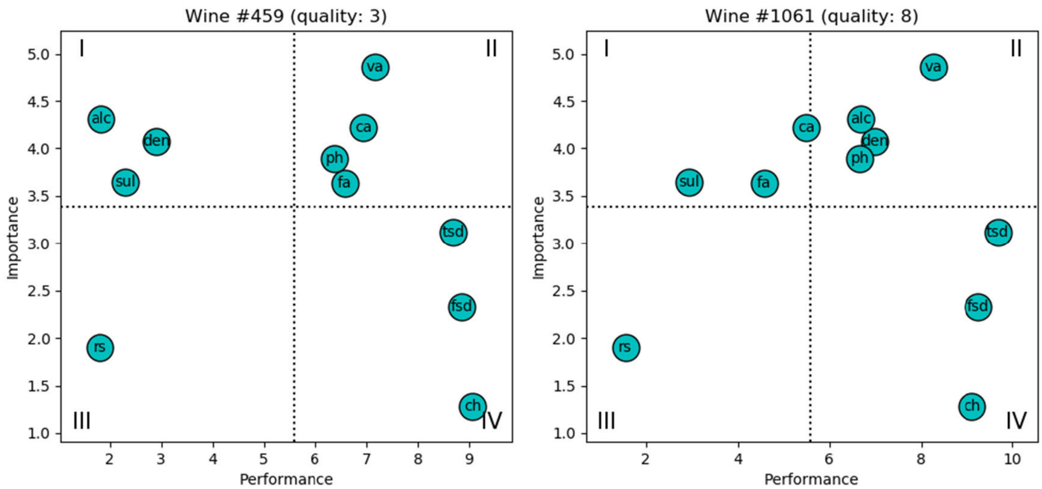

FIGURE 22. Importance performance analysis for two examples of red wines of low (left) and high (right) perceived quality.

TABLE 2. Quadrants of the importance performance analysis.

Quadrant	Meaning	Importance	Performance
I	Concentrate here	High	Low
II	Keep up the good work	High	High
III	Low priority	Low	Low
IV	Possible overkill	Low	High

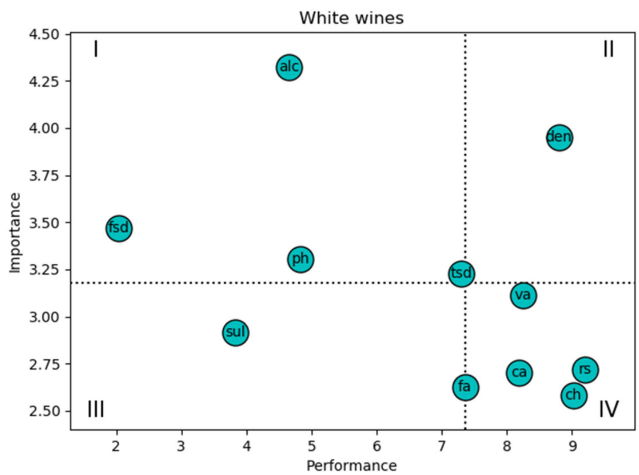

FIGURE 23. Importance performance analysis for the white wine dataset.

This quadrant is usually labelled as ''concentrate here'' in the IPA terminology. It indicates that the main physicochemical attributes of a red wine where to concentrate efforts should be, in an ordered list, citric acidity (*ca*), alcohol (*alc*), sulphates (*sul*) and fixed acidity (*fa*).

Similar analysis can be developed for a particular wine. Then, given a certain wine, an informed decision can be taken to improve the perceived quality of that wine. In this case, the local performance is employed instead of the global one. The IPA for two examples of low and high quality red wines are depicted in Fig. 22.

As for the white wine dataset, the global IPA is depicted in Fig. 23. It indicates that the main physicochemical attributes of a white wine where to concentrate efforts should be, in an ordered list, alcohol (*alc*), free sulfure dioxide (*fsd*) and *ph*. Alcohol is, therefore, a main vector of perceived quality for both red and white wines.

V. DISCUSSION

A. COMPARISON OF IMPORTANCE METRICS

In Section III-A, up to 8 importance metrics and their behavior on synthetic datasets have been presented. Then, in Section IV-A, these 8 importance metrics have been applied to the wine datasets. Finally, in Section IV-B, one of the importance metrics, that based on JSD, has been informally selected for its use in IPA. Now, a more detailed comparison of importance metric is undertaken.

For this purpose, the importance values are calculated when they are applied to the synthetic data sets (detailed in Section 0). Then several indicators, also defined in that Section, are obtained and compared. In Fig 24, the linearity coefficient and the statistical mean power for each importance metric are depicted. Firstly, it can be seen that importance based on mutual information shows a high mean power but a low linearity for most of the datasets.

Secondly, monotonic functions (linear, logarithmic and cubic) show a high linearity for most of the importance metrics. However, non-monotonic relationships are unproperly characterized by the simplest and most common methods: regression, Pearson, Spearman and Kendall. On the other hand, maximal correlation and Linfoot metrics obtain good results on non-monotonic functions but relatively fail on the multivalued vertical sine relationship. These results are compatible with other previous studies [53]. On the contrary,

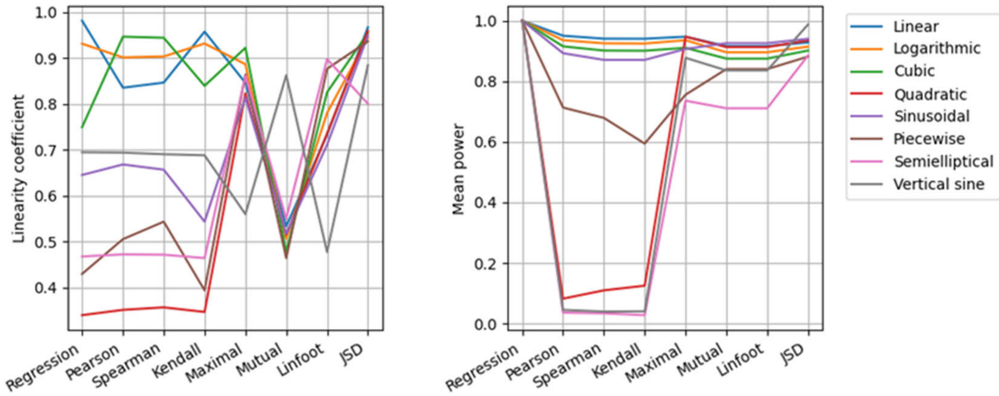

FIGURE 24. Linearity coefficient (left) and statistical mean power (right) of several importance metrics and relationships. These values have been obtained using the synthetic datasets.

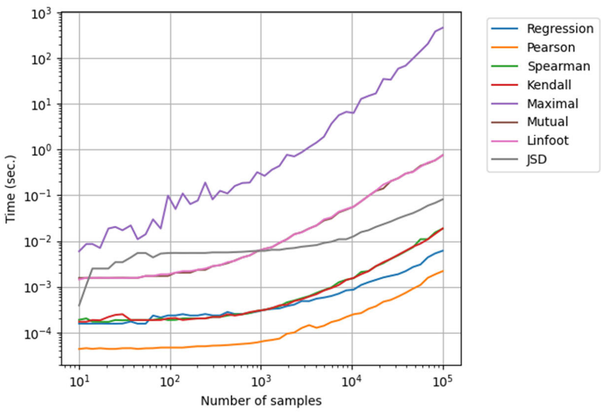

FIGURE 25. Time required to compute the importance metric as a function of the number of samples in the dataset. Values obtained on a standard desktop computer using a monotonic cubic relationship with a medium level of noise ($\beta = 0.5$).

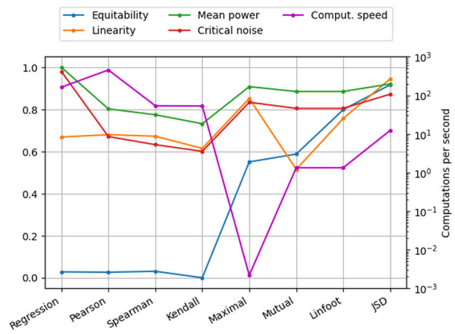

FIGURE 26. Comparison of different importance metrics when they are applied to the eight synthetic datasets (with 10^5 samples each one). Linearity coefficient, mean power and critical noise values have been averaged among the eight datasets.

JSD shows a very good statistical mean power for every relationship, while it has an excellent linearity for single-valued functions and a good linearity for multivalued dependances (semielliptical and vertical sine).

Another issue that should be discussed in the comparison of different importance metrics is the computational effort required to obtain them. The time required to compute any importance mainly depends on the number of samples in the dataset, as it is depicted in Fig. 25 (in logarithmic scale). These times have been obtained on a standard desktop computer by calculating the importance metrics of a monotonic cubic relationship, with a medium level of noise ($\beta = 0.5$).

It can be observed that Pearson has the lowest computing times. The remaining most common methods (regression, Spearman, and Kendall) show computing times about an order of magnitude higher. On the other hand, importance metrics based on information theory (mutual, Linfoot and JSD) require a computing effort about two orders of magnitude higher than Pearson. Finally, maximal correlation method demands a huge computing effort (about four orders of magnitude over Pearson) limiting their use in many situations.

A summary of the performance obtained by every importance metric is depicted in Fig 26. Firstly, equitability is shown which, by definition, has a single value for each importance metric. Secondly, the values of linearity coefficient, mean power and critical noises have been averaged from those obtained for every synthetic dataset. And finally, computer speed is obtained as the inverse of the computation time for a dataset with 10^5 samples and a cubic relationship between quality and feature.

As for equitability, it presents a very low value for classical metrics (regression, Pearson, Spearman and Kendall) as they fail to seize non-monotonic relationships. JSD excels on equitability, with Linfoot obtaining slightly lower values. On the other hand, linearity coefficient shows intermediate values for almost every importance metric, as it is an average for several synthetic datasets. However, JSD stands out as it has a very high linearity for all the analyzed datasets.

Statistical power of the importance metrics is measured using two indicators: mean power and critical noise. Both offer similar results with excellent results for regression, closely followed by JSD. Finally, computing speed is high

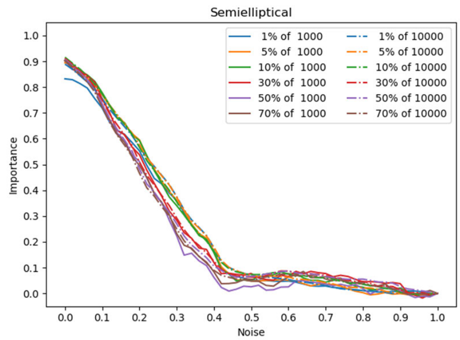

FIGURE 27. Importance based on the Jensen-Shannon Divergence (JSD) for the semielliptical synthetic datasets as a function of the noise level (β). Continuous lines correspond to datasets with 1000 samples; dashed lines correspond to 10000 samples. Each line represents the estimation of JSD for a certain a number of bins (as a percentage of the dataset size).

for classical methods, medium for information-based metrics, and very low for maximal correlation.

According to this analysis, the use of JSD in Section 0 for wine datasets is fully justified, as it shows the best equitability, the best linearity, and the second best statistical power while it requires a medium and affordable computational effort.

All the importance metrics described and compared through the article can be defined as model-free, in the sense that they do not rely on any underlying model (neither algorithmic nor machine learning) of the relationship between properties and quality (the only exception is the importance based on regression where a very simple linear model is assumed, as expressed in Eq. 5).

Alternatively, when a machine learning model is trained for any other purpose (i.e. prediction of quality for a new wine), some additional importance metrics wrapped to that model can be defined, such as SHAP [54], Lime [55] or Tree Importance Scores [56]. However, these metrics, as *a posteriori* to the training, are better interpreted as an explanation of the model predictions. Furthermore, they can lead to different importance values when several machine learning algorithms are considered and, additionally, more computing resources are required due to the previous training process. For these reasons, the model-wrapped importance metrics have not been considered in this research although they could be contemplated in further developments.

The wine dataset described in Section 0 has been used as an example of application of the methodologies proposed in this research. In the case of using a different wine dataset, where some context variables differ (e.g., vineyard location, temperature, humidity, rainfall, soil composition …) or other physicochemical properties (features) of the wine were measured, the same procedure to obtain each importance feature should be employed. Obviously, in these cases, different results could be obtained.

The importance metrics studied in this research consider the effect of individual physicochemical properties (features) on the perceived quality of the wine. However, the features are not independent, and their values can be correlated (for example, fixed, volatile, and citric acidity are correlated with each other and with pH). The combined impact of several features (even the full set of properties) on the quality would require an expanded definition of the importance metrics, which deserves more detailed analysis in future developments.

Although the quality of the wine has been the target variable in this study, some other subjective characteristics can be analyzed such as, for example, sweetness, balance, harmony, complexity, precision, elegance, power, etc. These subjective perceptions, in addition to quality, can be used as a basis for commercial strategies, which must also consider some other aspects such as price, bottle design, labeling, etc. These commercial aspects go beyond the scope of the present work.

B. BIN SIZE FOR JSD

As it has been explained in Section III-C, computing Jensen-Shannon Divergence requires to estimate the probabilities of the feature values for every value of the quality. To obtain these probabilities, the full range of the feature values $[x_{min}, x_{max}]$ is divided into a certain number of bins, n_b. Then, the k-th bin has a central value x_k, while it contains a number of samples, n_k. If the total number of samples is n, the (density of) probability that $x = x_k$ (be in the k-th bin) is $p_k = n_k/n$. As any other bin-based method, the value of JSD may depend on the number of bins used and the number of instances in the dataset. Fortunately, it has been observed in the synthetic dataset that these parameters have very little influence on the value of JSD. As an example, Fig. 27 shows the evolution of JSD for one of the most problematic functions (semielliptical relationship) when two sizes of the dataset and various number of bins (as a percentage of the dataset size) are used. Similar results are obtained for other relationships.

C. JSD FOR DISCRETE MULTIVALUED FUNCTIONS

In Section III) the theoretical importance value for noiseless relationships was corrected for discrete multivalued functions. To show the behavior of JSD when faced with this type of difficult relationships, several multivalued functions, with five different shapes, have been synthetized as it is depicted in Fig 28. Then for each of these shapes, the number of quality values for each feature value is changed. For instance, the first column in Fig 28 shows a linear multivalued function with $m = 4$ (top) and $m = 2.5$ (bottom) values of quality for each feature value.

Then, the quality values are discretized using L_q levels ($L_q = 11$ has been used in the previous sections). Finally, for all the noiseless functions, the JSD is obtained as a function of L_q, and the number of quality values for each feature value, m. The result is depicted in Fig 29. The theoretical value derived in Section III) is also shown as a dotted line.

It can be observed that single-valued functions ($m = 1$) obtains an importance $I_j \approx 1$. However, in the case of multivalued functions, the importance for the noiseless linearly decreases with m and L_q, following a behavior very

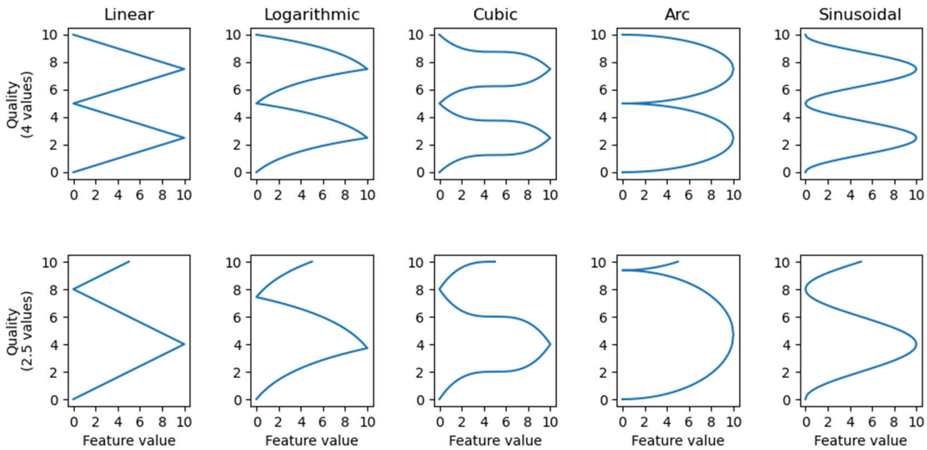

FIGURE 28. Examples of noiseless multivalued functions with m=4 (top) and m=2.5 (bottom) values of quality for each feature value.

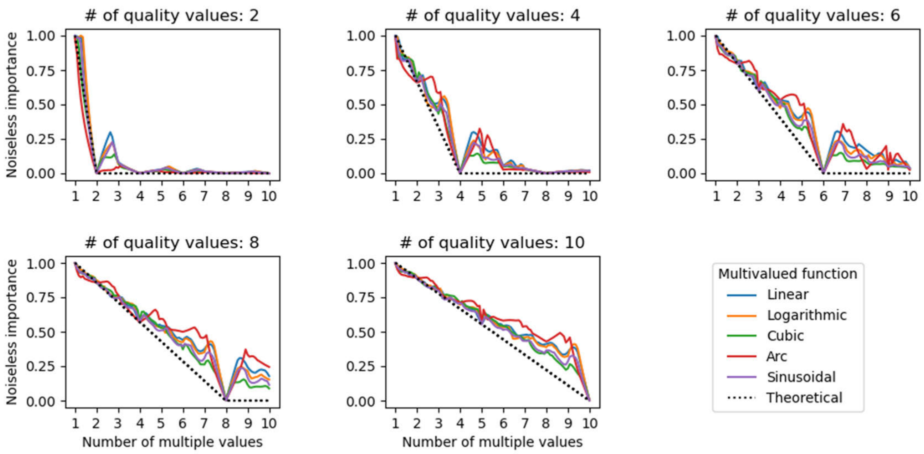

FIGURE 29. Importance of five noiseless relationships as a function of the number of quality values for each feature value. Each plot is obtained for the quality discretized in different number of levels.

similar to the theoretical model. Then, it can be asserted that JSD properly reflects the importance of discrete multivalued functions.

VI. CONCLUSION

Through the paper, a systematic comparison of different importance metrics has been undertaken using eight synthetic qualitatively different relationships with numerous levels of randomness (noise). For that comparison five measurable indicators have been employed, three of them proposed in previous works (equitability, statistical power and computation speed) and two novel indicators: linearity coefficient and critical noise.

The cases of relationships with a discrete-valued target variable (as the perceived quality of wines) have been extensively studied. For them, a novel importance metric based on JSD has been introduced. Additionally, the usual concept of importance has been adapted to properly reflect the multivalued functions and a novel correction factor has been introduced.

In the synthetic datasets explored in this research, the JSD-based importance metric clearly overperforms other metrics previously proposed in the literature, at a very moderate and affordable computational cost. Also, it can be asserted that JSD properly reflects the importance of discrete multivalued functions.

The use of eight importance metrics to describe the importance of each feature in the wine datasets has been explored. Then these metrics are combined using PCA, obtaining a novel 2D map of features which can be used to prescribe the physicochemical properties to be addressed in the design and production process.

Finally, the JSD-based metric has been applied as the key element of an updated importance performance analysis in the wine datasets. This IPA has indicated that the main physicochemical attributes of a red wine where to concentrate efforts should be, in an ordered list, citric acidity (ca), alcohol (alc), sulphates (sul) and fixed acidity (fa). As for the white wine dataset, the main attributes are alcohol (alc), free sulfure dioxide (fsd) and ph. Alcohol is, therefore, a main vector of perceived quality for both red and white wines.

Identifying the physicochemical properties of wines with the greatest impact on perceived quality is a powerful tool to tune the winemaking process with the aim of improving perceived quality. This conclusion is not limited to the datasets analyzed in the paper but can also be extended to different wine production processes and many other engineering applications.

The main limitation of the proposed JSD-based importance metric is that it requires a discrete target variable (wine quality in this case). For its application to continuous target variables, they should be quantized previously, which could introduce some distortion which should be analyzed.

Future research developing the methodologies presented though this article could focus, first, on the impact of non-independent (correlated) features, extending the definition of importance to these cases and studying the combined effect of several correlated features (physicochemical properties) on the target variable (quality). This work can also be extended by considering and comparing importance metrics wrapped to a certain machine learning model. Also, commercially focused analysis of the importance metric and its application to other food and beverage products or other engineering fields deserve further investigation.

REFERENCES

[1] E. Pomarici, "Recent trends in the international wine market and arising research questions," *Wine Econ. Policy*, vol. 5, no. 1, pp. 1–3, Jun. 2016.

[2] R. Sellers and V. Alampi-Sottini, "The influence of size on winery performance: Evidence from Italy," *Wine Econ. Policy*, vol. 5, no. 1, pp. 33–41, Jun. 2016.

[3] J. R. Ferrer-Lorenzo, M. T. Maza-Rubio, and S. Abella-Garcés, "'Business model and performance in the Spanish wine industry,'" *J. Wine Res.*, vol. 30, no. 1, pp. 31–47, Jan. 2019.

[4] E. Pomarici and R. Sardone, "EU wine policy in the framework of the CAP: Post-2020 challenges," *Agricult. Food Econ.*, vol. 8, no. 1, pp. 1–40, Dec. 2020.

[5] S. B. Swami, N. J. Thakor, and A. D. Divate, "Fruit wine production: A review," *J. Food Res. Technol.*, vol. 2, no. 3, pp. 93–100, 2014.

[6] A. Artero, A. Artero, J. J. Tarín, and A. Cano, "The impact of moderate wine consumption on health," *Maturitas*, vol. 80, no. 1, pp. 3–13, Jan. 2015.

[7] L. Castaldo, A. Narváez, L. Izzo, G. Graziani, A. Gaspari, G. Di Minno, and A. Ritieni, "Red wine consumption and cardiovascular health," *Molecules*, vol. 24, no. 19, p. 3626, Oct. 2019.

[8] J. Varriano, *Wine: A Cultural History*. London, U.K.: Reaktion Books, 2022.

[9] D. K. Wright, H. Yoon, A. M. Morrison, and T. Šegota, "Drinking in style? Literature review of luxury wine consumption," *Brit. Food J.*, vol. 125, no. 2, pp. 679–695, Jan. 2023.

[10] S. Charters and S. Pettigrew, "The dimensions of wine quality," *Food Quality Preference*, vol. 18, no. 7, pp. 997–1007, Oct. 2007.

[11] A. J. V. Jover, F. J. Lloréns Montes, and M. D. M. Fuentes Fuentes, "Measuring perceptions of quality in food products: The case of red wine," *Food Quality Preference*, vol. 15, no. 5, pp. 453–469, Jul. 2004.

[12] P. Oyinseye, A. Suárez, E. Saldaña, P. Fernández-Zurbano, D. Valentin, and M.-P. Sáenz-Navajas, "Multidimensional representation of wine drinking experience: Effects of the level of consumers' expertise and involvement," *Food Quality Preference*, vol. 98, Jun. 2022, Art. no. 104536.

[13] R. A. de Wijk and L. P. J. J. Noldus, "Using implicit rather than explicit measures of emotions," *Food Quality Preference*, vol. 92, Sep. 2021, Art. no. 104125.

[14] S. Lagast, X. Gellynck, J. J. Schouteten, V. De Herdt, and H. De Steur, "Consumers' emotions elicited by food: A systematic review of explicit and implicit methods," *Trends Food Sci. Technol.*, vol. 69, pp. 172–189, Nov. 2017.

[15] E. Pittari, L. Moio, and P. Piombino, "Interactions between polyphenols and volatile compounds in wine: A literature review on physicochemical and sensory insights," *Appl. Sci.*, vol. 11, no. 3, p. 1157, Jan. 2021.

[16] M. Mora, A. D. de Matos, L. Vázquez-Araújo, V. Puente, J. Hernando, and C. Chaya, "Exploring young consumers' attitudes and emotions to sensory and physicochemical properties of different red wines," *Food Res. Int.*, vol. 143, May 2021, Art. no. 110303.

[17] J. Yang and J. Lee, "Current research related to wine sensory perception since 2010," *Beverages*, vol. 6, no. 3, p. 47, Jul. 2020.

[18] S. E. Ebeler, "Analysis of grapes and wines: An overview of new approaches and analytical tools," in *Proc. ACS Symp. Ser.*, vol. 1203, Nov. 2015, pp. 3–12.

[19] S. Poornima and M. Pushpalatha, "A survey on various applications of prescriptive analytics," *Int. J. Intell. Netw.*, vol. 1, pp. 76–84, Jan. 2020.

[20] A. Luque, A. De Las Heras, M. J. Ávila-Gutiérrez, and F. Zamora-Polo, "ADAPTS: An intelligent sustainable conceptual framework for engineering projects," *Sensors*, vol. 20, no. 6, p. 1553, Mar. 2020, doi: 10.3390/S20061553.

[21] E. S. King, R. L. Dunn, and H. Heymann, "The influence of alcohol on the sensory perception of red wines," *Food Quality Preference*, vol. 28, no. 1, pp. 235–243, Apr. 2013.

[22] K. Pearson, "Notes on the history of correlation," *Biometrika*, vol. 13, no. 1, pp. 25–45, 1920.

[23] M. Bermudez-Edo, P. Barnaghi, and K. Moessner, "Analysing real world data streams with spatio-temporal correlations: Entropy vs. Pearson correlation," *Autom. Construct.*, vol. 88, pp. 87–100, Apr. 2018.

[24] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 145–151, Jan. 1991.

[25] J. A. Martilla and J. C. James, "Importance-performance analysis," *J. Mark*, vol. 41, no. 1, pp. 77–79, Jan. 1977.

[26] J. Abalo, J. Varela, and V. Manzano, "Importance values for importance–performance analysis: A formula for spreading out values derived from preference rankings," *J. Bus. Res.*, vol. 60, no. 2, pp. 115–121, Feb. 2007.

[27] A. K. Chee, "The mechanistic determination of doping contrast from Fermi level pinned surfaces in the scanning electron microscope using energy-filtered imaging and calculated potential distributions," *Microsc. Microanal.*, vol. 28, no. 5, pp. 1538–1549, Oct. 2022.

[28] A. K. W. Chee, "Enhancing doping contrast and optimising quantification in the scanning electron microscope by surface treatment and Fermi level pinning," *Sci. Rep.*, vol. 8, no. 1, p. 5247, Mar. 2018.

[29] A. K. W. Chee, "Quantitative dopant profiling by energy filtering in the scanning electron microscope," *IEEE Trans. Device Mater. Rel.*, vol. 16, no. 2, pp. 138–148, Jun. 2016.

[30] A. K. W. Chee, R. F. Broom, C. J. Humphreys, and E. G. T. Bosch, "A quantitative model for doping contrast in the scanning electron microscope using calculated potential distributions and Monte Carlo simulations," *J. Appl. Phys.*, vol. 109, no. 1, pp. 1–9, Jan. 2011, doi: 10.1063/1.3524186.

[31] D. Dua and C. Graff. (2017). *UCI Machine Learning Repository*. [Online]. Available: http://archive.ics.uci.edu/ml

[32] P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis, "Modeling wine preferences by data mining from physicochemical properties," *Decis. Support Syst.*, vol. 47, no. 4, pp. 547–553, Nov. 2009.

[33] T. Alamo, A. Luque, D. R. Ramirez, and R. Tempo, "Randomized control design through probabilistic validation," in *Proc. Amer. Control Conf. (ACC)*. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, Jun. 2012, pp. 839–844, doi: 10.1109/ACC.2012.6314838.

[34] M. Pal and P. Bharati, "Introduction to correlation and linear regression analysis," in *Applications of Regression Techniques*. Singapore: Springer, 2019, pp. 1–18.

[35] K. R. Murphy, B. Myors, and A. Wolach, *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*. Evanston, IL, USA: Routledge, 2014.

[36] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti, "Detecting novel associations in large data sets," *Science*, vol. 334, no. 6062, pp. 1518–1524, Dec. 2011.

[37] C. Spearman, "The proof and measurement of association between two things," *Amer. J. Psychol.*, vol. 15, no. 1, pp. 72–101, 1904.

[38] N. S. Chok, "Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data," Ph.D. dissertation, Univ. Pittsburgh, Pittsburgh, PA, USA, 2010.

[39] L. M. Adler, "A modification of Kendall's tau for the case of arbitrary ties in both rankings," *J. Amer. Stat. Assoc.*, vol. 52, no. 277, pp. 33–35, Mar. 1957.

[40] M. G. Kendall and J. D. Gibbons, *Rank Correlation Methods*, J. D. Gibbons, Ed., 5th ed. London, U.K.: Edward Arnold, 1990.

[41] S. Asoodeh, F. Alajaji, and T. Linder, "On maximal correlation, mutual information and data privacy," in *Proc. IEEE 14th Can. Workshop Inf. Theory (CWIT)*, Jul. 2015, pp. 27–31.

[42] N. Papadatos and T. Xifara, "A simple method for obtaining the maximal correlation coefficient and related characterizations," *J. Multivariate Anal.*, vol. 118, pp. 102–114, Jul. 2013.

[43] D. Wang and M. Murphy, "Estimating optimal transformations for multiple regression using the ACE algorithm," *J. Data Sci.*, vol. 2, no. 4, pp. 329–346, Jul. 2021.

[44] S. Watanabe, "Information theoretical analysis of multivariate correlation," *IBM J. Res. Develop.*, vol. 4, no. 1, pp. 66–82, Jan. 1960.

[45] A. Dionisio, R. Menezes, and D. A. Mendes, "Mutual information: A measure of dependency for nonlinear time series," *Phys. A, Stat. Mech. Appl.*, vol. 344, nos. 1–2, pp. 326–329, Dec. 2004.

[46] J. B. Kinney and G. S. Atwal, "Equitability, mutual information, and the maximal information coefficient," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 9, pp. 3354–3359, Mar. 2014.

[47] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 6, p. 66138, Jun. 2004.

[48] B. C. Ross, "Mutual information between discrete and continuous data sets," *PLoS ONE*, vol. 9, no. 2, Feb. 2014, Art. no. e87357.

[49] E. H. Linfoot, "An informational measure of correlation," *Inf. Control*, vol. 1, no. 1, pp. 85–89, Sep. 1957.

[50] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.

[51] F. L. Gewers et al., "Principal component analysis: A natural approach to data exploration," *ACM Comput. Surveys (CSUR)*, vol. 54, no. 4, pp. 1–34, 2021.

[52] J. C. Gower, S. Lubbe, and N. J. Le Roux, *Understanding Biplots*. Hoboken, NJ, USA: Wiley, 2011.

[53] O. Rainio, "Different coefficients for studying dependence," *Sankhya B*, vol. 84, no. 2, pp. 895–914, Nov. 2022.

[54] S. M. Lundberg, P. G. Allen, and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–10.

[55] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[56] W. Loh, "Fifty years of classification and regression trees," *Int. Stat. Rev.*, vol. 82, no. 3, pp. 329–348, Dec. 2014, doi: 10.1111/INSR.12016.

AMALIA LUQUE (Member, IEEE) was born in Seville, Spain, in 1982. She received the degree in industrial engineering, in 2007, the master's degree in automation, robotics and telematics, in 2010, and the Ph.D. degree in industrial engineering, in 2014. She has been involved in teaching related to project engineering with the University of Seville, since 2015, where she is currently a Professor. Her research interests include control, business intelligence, data mining, feature extraction, and artificial intelligence.

MIRKO MAZZOLENI (Member, IEEE) was born in Ponte San Pietro, Italy, in 1989. He received the M.Sc. degree (summa cum laude) in computer engineering and the Ph.D. degree in engineering and applied sciences (control systems) from the University of Bergamo, Italy, in 2014 and 2018, respectively. He is currently an Assistant Professor with the University of Bergamo. His main research interests include system identification, machine learning, and fault diagnosis.

FRANCISCO ZAMORA-POLO was born in Badajoz, Spain, in 1981. He received the degree in industrial engineering from the University of Extremadura, Badajoz, in 2005, and the Ph.D. degree in industrial and graphic science and technology from the University of Extremadura, in 2011.

From 2007 to 2019, he was an Assistant Professor in mechanical engineering with the University of Extremadura. He is currently an Assistant Professor in project engineering with the University of Seville, Spain. His main research interests include business intelligence, data mining, feature extraction, and artificial intelligence.

Dr. Zamora-Polo is a member of IPMA, Spain.

ANTONIO FERRAMOSCA (Member, IEEE) was born in Maglie (LE), Italy, in 1982. He is an Associate Professor at the Department of Management, Information and Production Engineering of the University of Bergamo (Italy). He received his Bachelor's and Master's degrees in Computer Science Engineering, both from the University of Pavia (Italy), in 2004 and 2006 respectively, and the Ph.D. degree in Engineering, from the University of Seville (Spain) in 2011. He was a Postdoctoral Fellow and then Research Associate at the Argentine National Scientific and Technical Research Council (CONICET), (from 2012 to 2020). He was a visiting researcher at the Department of Chemical and Biological Engineering at University of Wisconsin-Madison (USA) in 2009, and a visiting professor at the Department of Automation and Systems Engineering at the University of Seville (Spain) in 2015. He serves as Associate Editor for the Optimal Control Applications and Methods journal, and for the Journal of Control, Automation and Electrical Systems. He is the author and co-author of more than 100 publications. His research interests are in the field of control theory for constrained linear and nonlinear systems, including Model Predictive Control, learning-based control, Artificial Pancreas, and control applications to industrial and biological systems.

JUAN RAMÓN LAMA was born in Seville, Spain, in 1969. He received the degree in electronic engineering, in 1992, the master's degree in electronic engineering, in 2006, and the Ph.D. degree in industrial engineering, in 2016. He has been involved in teaching related to the project engineering with the University of Seville, since 1996, where he is currently an Associate Professor. His research interests include advanced manufacturing systems and intelligent systems for quality management in extended enterprise environments and, more recently, sustainable industrial systems.

FABIO PREVIDI (Member, IEEE) was born in Milan, Italy, in 1968. He received the M.Sc. degree in electronic engineering (control systems) and the Ph.D. degree in computer science and control systems engineering from Politecnico di Milano, Milan, in 1993 and 1999, respectively. From 1999 to 2000, he was a Research Assistant with the Centre for Systems and Control, Department of Mechanical Engineering, University of Glasgow, Glasgow, U.K. From 2000 to 2002, he was a Research Assistant with Dipartimento di Elettronica e Informazione, Politecnico di Milano. From 2002 to 2006, he was an Assistant Professor with the Engineering Faculty, Universitá degli Studi di Bergamo, Dalmine, Italy, where he became an Associate Professor, in 2006, and a Full Professor, in 2017. His research interests include fault diagnosis, system identification, control of biomedical systems, and application of control systems methods in physics.