Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE 7 (2020) 342–354

journal homepage: www.elsevier.com/locate/inpa

Multi-target prediction of wheat flour quality parameters with near infrared spectroscopy

Sylvio Barbon Junior [a,*], *Saulo Martielo Mastelini* [a], *Ana Paula A.C. Barbon* [b], *Douglas Fernandes Barbin* [c], *Rosalba Calvini* [d], *Jessica Fernandes Lopes* [a], *Alessandro Ulrici* [d]

[a] *Computer Science Department, Londrina State University (UEL), Londrina 86057-970, Brazil*
[b] *Animal Science Department, Londrina State University (UEL), Londrina 86057-970, Brazil*
[c] *Department of Food Engineering, Campinas State University (UNICAMP), Campinas 13083-862, Brazil*
[d] *Department of Life Science, University of Modena and Reggio Emilia (UNIMORE), 42122 Reggio Emilia, Italy*

ARTICLE INFO

Article history:
Received 12 November 2018
Received in revised form
4 June 2019
Accepted 9 July 2019
Available online 12 July 2019

Keywords:
Random forest
Support Vector Machine
Near-infrared spectroscopy
Machine learning
MTRS
DSTARS
Partial Least Squares
ERC

ABSTRACT

Near Infrared (NIR) spectroscopy is an analytical technology widely used for the non-destructive characterisation of organic samples, considering both qualitative and quantitative attributes. In the present study, the combination of Multi-target (MT) prediction approaches and Machine Learning algorithms has been evaluated as an effective strategy to improve prediction performances of NIR data from wheat flour samples. Three different Multi-target approaches have been tested: Multi-target Regressor Stacking (MTRS), Ensemble of Regressor Chains (ERC) and Deep Structure for Tracking Asynchronous Regressor Stack (DSTARS). Each one of these techniques has been tested with different regression methods: Support Vector Machine (SVM), Random Forest (RF) and Linear Regression (LR), on a dataset composed of NIR spectra of bread wheat flours for the prediction of quality-related parameters. By combining all MT techniques and predictors, we obtained an improvement up to 7% in predictive performance, compared with the corresponding Single-target (ST) approaches. The results support the potential advantage of MT techniques over ST techniques for analysing NIR spectra.

1. Introduction

Near-infrared (NIR) spectroscopy is a consolidated technology widely applied in several research fields such as food science [3,2,51,45,4,33,43,39,23,39], agriculture [47,30,42,13,17] and biomedicine [27,18]. The main advantages of NIR spectroscopy include the possibility of performing rapid, minimally invasive, simple, reagent-free and non-destructive measurements [33,47,42,20]. In this way, NIR analysis is capable of coping with the modernisation of processing industries, improve health diagnoses, reduce costs towards a sustainable alternative in food and raw material characterisation [32,18,41,42].

Essentially, NIR spectroscopy is based on absorption bands derived from overtones and combinations of fundamental vibrations of chemical bonds (mainly C—H, N—H, O—H, S—H,

* Corresponding author.
E-mail addresses: barbon@uel.br (S. Barbon Junior), mastelini@uel.br (S.M. Mastelini), dfbarbin@unicamp.br (D.F. Barbin), rosalba.calvini@unimore.it (R. Calvini), jessicafernandes@uel.br (J.F. Lopes), alessandro.ulrici@unimore.it (A. Ulrici).
Peer review under responsibility of China Agricultural University.
https://doi.org/10.1016/j.inpa.2019.07.001

C—C and C=O) observed in the 780–2500 nm range of the electromagnetic spectrum [40]. In general, a multivariate statistical approach is mandatory in order to extract the useful information from NIR spectra. In this context, multivariate analysis allows to relate the spectral properties of a representative set of samples with the desired response and to use the resulting model to predict new samples [22,8].

Usually, PLS is the multivariate approach more commonly applied for regression tasks [42]. [33] have highlighted that PLS is the linear-based method more often used for regression models in NIR spectra analysis [33].

In particular, multivariate linear methods (such as PLS) can be considered as a branch of Machine Learning (ML) supervised approaches, which are able to automatically detect patterns in the data and use these patterns to predict future data. However, in some situations changes in the physical and chemical properties of the analysed samples can cause deviations from linearity. Possible non-linear contributions may result from changes in temperature, particle size, viscosity and chemical composition of the samples [47]. In order to solve this issue, non-linear ML methods have been successfully applied to model NIR spectra in several scenarios [5,16,47].

Among non-linear ML algorithms, Support Vector Machines (SVM) has been proven to be an accurate and reliable method, often superior to other regression or classification methods in the field of spectroscopic analysis [45]. [42] described the improvements obtained with SVM compared to PLS. However, the same paper also highlighted that generally, SVM requires a strong optimisation of hyper-parameters and a feature selection step in order to obtain successful results.

Other ML techniques, such as Artificial Neural Networks (ANN), are able to handle non-linear features with a little advantage in prediction accuracy. [30] compared ANN and PLS in the prediction of organic carbon, pH and clay content of soils. The results showed similar performances of both methods, but some laboratory independent validation and on-line evaluation proved the over-performance of ANN models. An important drawback in the use of ANN was highlighted by [51]. Indeed, due to the high number of tunable hyper-parameters and to the limited number of samples usually available in real applications, ANN is hindered by overfitting problem.

In addition, a drawback of non-linear methods is the lack of interpretability. As a matter of fact, the relevance of the different spectral features is a very valuable information for the chemical interpretation of a regression or classification model. Considering model interpretability, [33], reported that linear based machine learning methods are superior to non-linear ones.

In this context, Random Forest (RF) algorithm can combine both model performances and interpretability, since it generally leads to highly accurate prediction and it can be helpful in spectra interpretation by RF importance. RF was proposed by [10] as a combination of decision tree classifiers in an ensemble. The author describes RF as a trustful method that always converges, avoiding overfitting problem. In the most common RF algorithm, split selection for decision tree building is performed based on the decrease of Gini impurity. This feature value, named as RF importance, provides a relative ranking of the spectral features as described in [38]. Therefore, RF is a useful tool for regression studies and it has a potential for modelling linear and non-linear spectral responses [24]. [6] explored eight different ML algorithms for the prediction of pork meat storage time, with RF achieving the best results.

Although ML non-linear algorithms have been proved to be profitable for spectroscopy, PLS is still a common strategy for modelling spectral data thanks to the ability of dealing with high dimensional and multicollinear data, and the possibility of identifying the relevant predictors with a gain in model interpretability [47,26]. Furthermore, spectroscopic data usually consist of numerous features (wavelengths) and relatively few samples. In this situation, the optimisation of many meta-parameters can easily lead to overfitting [31]. In order to combine the advantages of PLS and ML non-linear algorithms, PLS can be used as a sort of data pre-processing and data compression method to extract the relevant features from spectroscopic data, which can later be used as predictors for non-linear regression methods.

Another issue of great relevance is that, in practical applications, NIR spectroscopy can be used to predict multiple quality parameters of a given sample. In this situation, Multi-target (MT) approaches of Machine Learning (ML) allow to obtain better performances compared to traditional single-target modelling (ST) [43]. Indeed, MT methods gain advantages from the exploration of inter-target influences and allow to reduce overfitting [9,36,35]. Furthermore, a MT model provides a global comprehension of a given problem by also considering the relationships between the different targets, in addition to the relationships between features and expected predictions (targets) [9]. [37] highlighted that MT prediction has the ability to generate models representing a broad variety of real-world applications, from natural language processing to bioinformatics.

In this context, we propose the usage of MT techniques coupled with ML non-linear algorithms induced on Partial Least Squares regression (PLS) scores to predict statistically correlated targets from NIR spectral data. It is possible to transform NIR data into PLS scores in order to compress hundreds of variables into few relevant features. On the other hand, Multi-targets techniques coupled with base-learners from Machine Learning are capable of dealing with non-linear behaviour and noisy data with a little prediction error.

In the present work, different MT approaches have been tested: Multi-target Regressor Stacking (MTRS) [46], Ensemble of Regressor Chains (ERC) [46] and Deep Structure for Tracking Asynchronous Regressor Stack techniques from Multi-target (DSTARS) [36]. Each MT technique has been coupled with Support Vector Machine (SVM), Random Forest (RF) and Linear Regression (LR) as base-learners to perform non-linear and linear modelling, respectively. As a case study, the proposed MT approaches have been tested in the prediction of quality-related parameters of bread wheat [21].

Results exposed the advantage of Multi-target strategy reaching an improvement of 7% with ERC and RF. Independent of ML algorithm, in all cases, the usage of MT coupled with SVM or RF increased the predictive performance.

Predictions obtained from RF overcome the SVM performance. Specifically, the MTRS method achieved inferior results when coupled with an LR.

2. Materials and methods

2.1. NIR datasets

The bread wheat samples considered in this study were collected from experimental fields located in different Italian regions and derived from two subsequent harvesting years. On the whole, 391 bread wheat white flour samples were analysed with a Bruker MPA Multi Purpose FT-NIR Analyzer, equipped with an integrating sphere (reflectance mode) and a RT-PbS detector in the 12,500–3600 cm^{-1} range.

The spectra were acquired following the acquisition parameters defined in [19] (8 cm^{-1} resolution, 155 sample scans). For each white flour sample, four replicate measurements were performed in different days, each time considering a different 50.0 g aliquot. The sample aliquots were placed into a glass Petri dish and gently shaken to render the packing density as much uniform as possible. The FT-NIR instrument was equipped with a rotating device, in order to acquire an average signal over a relatively wide sample surface. Between the different measurement sessions, samples were stored in sealed plastic bags and put into a plastic box in a dark place at 4 °C. Finally, the spectrum of each sample was obtained as the average of the four replicate measurements.

For each sample, the following quality-related parameters (targets) have been determined using the corresponding reference methods: Hectolitre weight (*HlW*, kg hl^{-1}), Falling Number (*FN*, s), Protein content (*Prot*, % dry matter), Alveographic indexes (*W* and *P/L*), and Farinograph stability (*Stab*, min). Further details about the quality parameters can be found in [11].

The linear statistical correlation between the considered quality parameters (*HlW*, *FN*, *Prot*, *W*, *P/L*, and *Stab*) can be evaluated in Fig. 1, which reports the Pearson correlation coefficient calculated for the different parameters.

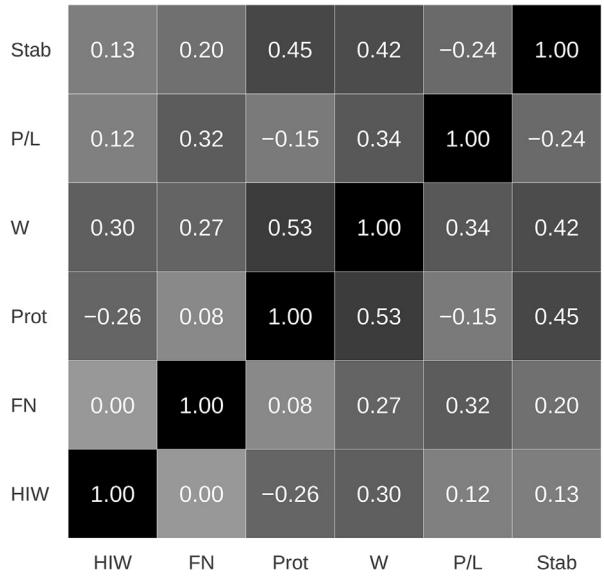

Fig. 1 – Pearson Correlation from targets of $Dataset_P$.

Kennard-Stone algorithm applied to the principal components space was used in order to split the whole dataset of NIR spectra into two datasets: the training cross-validation set ($Dataset_{CV}$) composed of 200 samples and the prediction set ($Dataset_P$) of 91 samples. The cross-validation set was used to calculate and optimise the regression models, while the prediction test set was employed to finally evaluate the predictive ability of the traditional ST approach compared to the multi-target ones. This subdivision of the dataset was performed in order to minimise the risk of overfitting.

2.2. Learn-based regressors

The training set of NIR spectra (obtained as reported in Section 2.1) has been firstly analysed by means of PLS. PLS is a multivariate analysis technique in which the independent variables are projected onto a small number of latent variables (LVs) to simplify the relationship between them and a predictive target [33]. The target variable is actively used in assessing the LVs to ensure that the first one is most relevant for predicting the targets. Usually, the relation between the scores extracted from LVs and the targets are built by a linear modelling toward prediction of novel samples. The PLS technique was implemented by SIMPLS algorithm [14]. In particular, the PLS scores have been extracted from the optimal PLS model (i.e., for the model with a number of LVs corresponding to the minimum value of the root mean square error in cross-validation), calculated on the mean-centred training set spectra for each one of the auto-scaled target variables. Subsequently, the test set spectra have been projected on the e PLS latent variable space and the corresponding scores have been calculated. These steps are highlighted in the overview of procedures as 1 and 2, reported in Fig. 2. In this manner, it was possible to extract the useful information contained in hundreds of spectral variables into few latent variables relevant for the prediction of the targets. The ST models were then calculated considering the PLS scores selected for each target, while the ensemble of all the PLS scores was used as input for the subsequent development of multi-target regression models. In particular, the three different MT approaches described in Section 2.3 have been tested (ERC, MTRS and DSTARS).

MT algorithms require a learn-based regressor in the stack or chain structure to induce each model. In order to evaluate the benefits induced by linear and non-linear learn-based regressors, MT approaches have been coupled with Linear Regression (LR), Support Vector Machine (SVM) and Random Forest (RF), as shown in Step *a* of Fig. 2.

SVM is a regression algorithm from kernel-based methods. It can be used for solving many types of problems, presenting high accuracy and ability to treat high-dimensional data. Through kernel space transformation, this technique has the flexibility to model diverse data sources [7], increasing the input dimensional space and data separability. In this work, the SVM implementation adopted was the `e1071 R` package with default parameters.

RF was initially created focusing on classification tasks. This algorithm consists of a collection of structured tree predictors, in which all random vectors are independent, identically distributed and each tree attributes a vote for the most

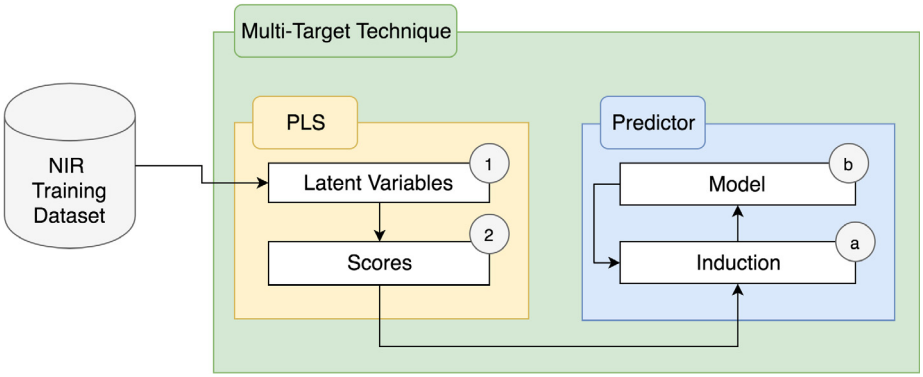

Fig. 2 – Overview of procedures performed.

popular class as the prediction. On the other hand, for regression purposes, the RF is composed of trees depending on a random vector, considering the average result over all trees in the Forest [10]. Each tree is grown using a bagging approach, where different training datasets are formed using bootstrap sampling. The `randomForest` R package was used with default parameters in experiments.

Therefore, each regression model has been adapted according to the requirements of the three MT techniques (Step *b* inside Fig. 2).

2.3. *Multi-target prediction*

The concept of Multi-target (MT) regression is related to the family of predictive problems with multiple continuous response variables, called as targets or outputs, following the assumption of being statistically correlated, i.e. the variation of a target has interference on the behaviour of the other responses [1,48,9,46]. Traditionally, MT tasks have been solved through two different approaches: Algorithm Adaptation and Problem Transformation [9].

Algorithm Adaptation, also named as multi-output adapted, is based on the adaptation of single target regression methods to handle multiple outputs exploring the dependencies among them with an enhancement in the predictive capability. This strategy is performed by modifying the original modelling method, for example changing the optimisation functions (e.g. SVMs) [34,52,50] or the node splitting criteria (e.g. Regression Trees) [29]. This type of approach has been successfully proposed in the last years focusing on different tasks [28,49,9,25,53].

In general, multi-output adapted models generate effective predictors by taking advantage of statistical target dependencies to create a single model. However, this approach could be more challenging since it aims not only to predict multiple targets but also to model and interpret their relationships at once [36]. [52] proposed the modification of task's input space through a virtualisation technique so that a MT task could be modelled as a wider single target problem. The authors used a Support Vector Regression (SVR) and achieved results comparable to ST strategy.

Another approach to address MT scenarios is Problem Transformation. This approach is based on calculating independent ST regression models for each target and finally concatenating the predictions. The calculation of several independent ST models to represent a single MT problem involves a strong increase of computational efforts and in several cases a loss of model interpretation by ignoring the dependencies between targets. On the other hand, this kind of approach offers considerable advantages. The first one is the possibility of applying any base-learner, or even more than one, toward better predictive performance and appropriate problem addressing. Also, adaptation methods improve the modularity and conceptual simplicity, with significantly better predictive performance than state-of-the-art methods [46].

Tsoumakas et al., in [48], proposed the use of random linear targets combinations to explore the relations between output values. This approach increases the original feature space dimension and solves multiple ST problems in the transformed space. The predicted values are used to solve a linear system and obtain the original targets predictions. In the last years, some MT methods have been modified and adapted from the area of multi-label classification [9,46]. Spyromitros-Xioufis et al. [46] proposed two techniques based on Problem Transformation: MTRS (Multi-target Regressor Stacking) and ERC (Ensemble of Regressor Chains). These two techniques are widely known and resulted to be adequate for a diversity of MT scenarios.

Santana et al. [44], proposed the Deep Regressor Stack, an idea similar to MTRS that consists in using targets approximations as additional predicting features in naive deep learning method. However, the drawback of huge memory requirements, as well as inappropriate dimensionality growth, preclude its usage associated with NIR spectra analysis. More recently, Santana et al., in a different work [43], proposed a new method, Multi-target Augmented Stacking (MTAS), addressing multi-target regression to predict twelve poultry meat characteristics. The authors explored the usage of Principal Components Analysis (PCA). Different from [43], in this work, we take advantage of PLS toward exploring DSTARS [36] MT approach. PLS compression leads to the extraction of features that are more relevant to target prediction.

Basically, DSTARS consists in a modification of MTRS built under the assumption that a model induction by deeper layers could offer better predictive performances than just one layer (ST) or two layers (MTRS). Therefore, DSTARS can take

advantage of different linear and non-linear relations among the targets, leading to possible improvements in the prediction performances of complex data, such as NIR spectra.

In the following sections a detailed description of MTRS, ERC and DSTARS algorithms is given.

2.3.1. *Multi-target Regressor Stack*

MTRS approach is based on additional input features from the output of ST models induction. In this way, considering a dataset composed by $X = \{x_1, x_2, \ldots, x_n\}$ input features and $Y = \{y_1, y_2, \ldots, y_d\}$ target variables, MTRS adds the $Y' = \{y'_1, y'_2, \ldots, y'_d\}$ from ST predictions as inputs, creating a new training dataset $X' = \{x_1, x_2, \ldots, x_n, y'_1, y'_2, \ldots, y'_d\}$. The new training dataset is used by each y to train another ST predictors layer, whose outputs are the final predictions.

New income input features are first merged into the first predictors' layer to obtain the output approximations. The approximations compose an augmented dataset toward forming the second level of predictors for the purpose of performing the final induction. Therefore, MTRS introduces inter-target relationships by the use of ST outputs into the modelling, increasing the prediction performance. The training procedure of MTRS is presented in Algorithm 1.

Algorithm 1 (*MTRS training algorithm*).

```
function MTRS(X, Y, d)
    Y' ← {}
    Level_0 ← {}
    // ST model induction
    for t = 1 to d do
        h: X → Y^t
        Y'^,t ← predict(h, X)
        Level_0 ← {Level_0, h}
    // Augmented training set definition
    X' ← X||Y'
    Level_1 ← {}
    for t = 1 to d do
        h: X' → Y^t
        Level_1 ← {Level_1, h}
    mtrs ← {Level_0, Level_1}
return mtrs
```

2.3.2. *Ensemble of Regressor Chains*

The ERC method consists in building a set of randomly ordered chains for each target to produce ST models through a generated sequence [46]. For each chain, initially, a ST model is induced using the first output prediction of the sequence. New models are then induced following the chain order, where each new regressor is trained over the augmented input dataset formed by the original input features and the previous models' predictions. This process is repeated until the end of all chain sequences. After training all models inside the chain, the predicted value of a novel sample is the average value obtained from chain's regressors.

The method prediction for a target y^t, $t = [1, d]$, is the average of the y^t predicted values over all chains. Since the output predictions are composed of values from different sorted chains, multiple levels of combinations and interdependence between targets are explored. ERC creates all possible target permutations if their number is less than 10 ($d \leq 3$), otherwise, the author suggests to choose ten random combinations.

The ERC's training step is presented in Algorithm 2. The *permute* procedure refers to a function which receives a set of elements and returns all their possible permutations. It is possible to perform all possible permutations without setting this parameter or to specify the desired number of combinations as an argument (in the original ERC's formulation, the mentioned parameter is equal to 10).

Algorithm 2 (*ERC training algorithm*).

```
function ERC(X, Y, d)
    targets ← names(Y)
    if d ≤ 3 then
        Chains ← permute(targets)
    else
        Chains ← permute(targets, 10)
    for chain in Chains do
        models_chain ← {}
        // To build augmented training sets
        X_aug ← X
    for t in chain do
        // ST model induction
        h : X_aug → Y^t
        y_pred ← predict(h, X_aug)
        // Extend training set with ST predictions
        X_aug ← X_aug || y_pred
        models_chain ← {models_chain, h}
        erc ← {erc, models_chain}
return erc
```

2.3.3. *Deep structure for Tracking Asynchronous Regressor Stack*

DSTARS [36] is based on the hypothesis that deeper layers could emphasise the relationships among targets which are statistically correlated. Instead of using additional features from only one single layer as ST estimators, or strictly two layers as MTRS, DSTARS sequentially builds new output predictions from as many layers as those necessary to minimise the error of a validation set. Iteratively, each best target prediction (evaluated using the validation set) is used as an additional predictive feature, augmenting the dataset. The procedure of the DSTARS algorithm can be split into two main steps: Tracking and Modelling. The former determines the best layer depth of targets variable, while the latter builds the final DSTARS model considering the whole modelling dataset.

The Tracking step, presented in Algorithm 3, starts with the subdivision of the dataset into training and validation sets. The authors recommend the use of a sampling approach, as the k-fold cross-validation, to increase the robustness. During the Tracking step, for each target new

models are calculated using the actual best approximations of all the targets as extra input variables. DSTARS keeps the tracking of the models that generated the best performance on layer l for each target y_t. Not all layers need to be used during final modelling for a particular target: in general, targets with low inter-correlations usually require a low layer depth, while some targets may require a greater layer depth. The final configuration is determined through a voting scheme, considering the reduction of the Root Mean Squared Error (RMSE) of the validation set brought by each layer and target combination. The tracking process ends when the decrease of the RMSE value obtained by the addition of a new predictors' layer $l+1$ is smaller than an ε parametrised value, leading to convergence.

Algorithm 3 (*DSTARS's tracking algorithm*).

```
function Tracking(X, Y, φ, ε, N_folds)
  // Dynamic regressor layer usage count
  Track ← [?, d]
  for k = 1 to N_folds do
    // Cross-validation data split
    {(X_tr, Y_tr), (X_val, Y_val)} ← CV(X, Y, k)
    errors [1..d] ← ∞
    converged [1..d] ← False
    layer ← 0
    T_tr ← T_val ← {}
    while !all(converged) do
      T_tr ' ← T_tr
      T_val ' ← T_val
      for t = 1 to d do
        // ST model induction
        mod : X_tr || T_tr → Y [t]_tr
        prd ← predict(mod, X_val || T_val)
        // Convergence stopping criteria
        if RMSE(Y [t]_val, prd) +ε < errors [t] then
          converged [t] ← False
          errors [t] ← RMSE(Y [t]_val, prd)
          // Error improvement counting
          Track[layer, t] ← Track[layer t] + 1
          T [t]_tr ' ← predict(mod, X_tr || T_tr)
          T [t]_val ' ← prd
        else
          converged [t] ← True
      T_tr ← T_tr '
      T_val ← T_val '
      layer ← layer + 1
  Track ← Track/N_folds // Layer normalisation
  // Threshold application: True or False values
  Track ← Track >φ
return Track
```

In Modelling step, a model h_t^l that outputs the t-th target estimation on the l-th layer is included in the final model only if it was used more than the φ percent of times during Tracking. New income samples are sequentially subjected to each layer of regressors and the last layer gives the predicted value for a specific target.

The usual DSTARS's structure starts by Tracking step and goes through the Modelling step presented in Algorithm 4.

Algorithm 4 (*DSTARS training algorithm*).

```
function DSTARS(X, Y, φ, ε, N_folds)
  T ← Tracking(X, Y, φ, ε, N_folds)
  dstars ← {}
  Y_approx ← {}
  for l = 1 to number of T rows do
    Y_aux ← Y_approx
    for t = 1 to d do
      if T [l, t] = True then
        // ST model induction
        mod : X || Y_approx → Y [t]
        Y [t]_aux ← predict(mod X || Y_approx)
        dstars ← {dstars, mod}
    Y_approx ← Y_aux
return dstars
```

2.4. Performance metrics

The performances of ST and MT methods have been evaluated on the $Dataset_{CV}$, sampled through a 10-fold cross-validation approach (CV), and the Prediction set ($Dataset_P$). In the final models, the whole Training set was used to build a single prediction model using the best MT hyperparameters determined in the CV approach, when considering the MTRS, ERC and DSTARS techniques.

Six different performance metrics were used to evaluate the calculated models: Mean Square Error (MSE), Average Relative Error (ARE), Average Root Mean Squared Error (aRMSE), Coefficient of Determination (R^2), average Relative Root Mean Square Error (aRRMSE), and Relative Performance (RP).

MSE corresponds to the mean squared difference between the predicted and the measured values. On the other hand, ARE was used to exposes the magnitude of the difference between the actual quality parameter and the prediction.

The aRMSE (Average Root Mean Square Error) is the mean of the Root Mean Square Error (RMSE) values obtained from each target. This last acts as a baseline and allows the measurement of the improvement over a shallow predictor. This metric has been used in various MT works [9] to compare non-homogeneous targets distributions. The aRRMSE, presented in Eq. (1), is computed averaging the d targets Relative Root Mean Square Error (RRMSE), and it was applied in this research to evaluate the improvement over ST models. The RRMSE for a given target t is the RMSE normalised by the average of the corresponding t.

$$aRRMSE = \frac{1}{d}\sum_{t=1}^{d}\sqrt{\frac{\sum_{k=1}^{N_{test}}\left(y_t^k-\hat{y}_t^k\right)^2}{\sum_{k=1}^{N_{test}}\left(y_t^k-\overline{y}_t\right)^2}} \tag{1}$$

The Coefficient of Determination (R^2) explains the proportion of the total variation associated with the dependent variable that is predictable from the model. The closer the R^2 values are to 1, the greater is the amount of variation of the dependent variable which is predictable by the regression model [12].

The Relative Performance (RP) compares for each target the aRRMSE of each MT method with the aRRMSE of the corresponding ST model. In this sense, it measures the increase (if RP > 1) or decrease (if RP < 1) in model performances [46]. The RP formulation is presented in Eq. (2).

$$RP_d(M) = \frac{aRRMSE(ST)}{aRRMSE(M)} \quad (2)$$

The aRRMSE supports the comparison of possible method superiority through the application of the Friedman's statistical test with significance level at $\alpha = 0.05$. The null hypothesis states that the performances of the MT methods are equivalent regarding the averaged aRRMSE per dataset. Any time the null hypothesis is rejected, the Nemenyi post hoc test can be applied, stating that the performance of two models is significantly different if the corresponding average ranks differ by at least a Critical Difference (CD) value. When multiple models are compared in this way, a graphic representation can be used to represent the results with the Critical Difference (CD) diagram, as previously proposed in [15].

Finally, the results obtained from ST and MT approaches and expressed with the error metrics cited above were analysed by means of Principal Component Analysis (PCA) in order to obtain a graphical overview of the whole model performances and evaluate the effectiveness of the different MT methods.

3. Results and discussion

3.1. Multi-target and Single-target comparison

In order to gain a first overview of the performance of the ST approach compared with MT, Table 1 reports the results obtained with linear regression coupled with Single-target and Multi-target methods. In this context, it must be highlighted that the linear regression of the PLS score values in the ST approach corresponds to the usual PLS regression. Table 1 shows that DSTARS and ERC resulted to be the best performing methods, with the lower aRRMSE value and the higher R^2 value obtained on the test set.

In general, the potential improvement induced by the different MT approaches is highlighted in Fig. 3, which reports the RP improvement for each algorithm (RF, SVM and LR). ERC led to improvements in all model performances obtaining the best average RP (1.015) from both data sets. The second best average RP was obtained by DSTARS (1.010); it did not improve the predictive performance only for RF in $Dataset_{CV}$. However, DSTARS overcame ERC results for LR algorithm in $Dataset_P$, obtaining 1.021 of RP. The worst results were achieved by MTRS with an average RP equals to 0.985. This approach was only able to contribute using SVM in $Dataset_{CV}$.

Concerning the different learn-based predictors, SVM took advantage from MT approaches with an increase of model performances, except for the case of the prediction of the test set ($Dataset_P$) coupled with MTRS. RF presented improvements with ERC, slightly boost coupled with DSTARS in $Dataset_P$, and no contributions over $Dataset_{CV}$. LR obtained the best RP with DSTARS, minor gains when combined with ERC and no improvements with MTRS.

In order to obtain a general overview of the performances of all the tested models, a PCA model was calculated on the results considering the metrics related to error performance from all the considered combinations of techniques and algorithms, and for both the CV and P datasets.

The optimal number of principal components (PC) has been found to be equal to 3, retaining the 97.72% of the total variance. Considering the behaviour of the loadings, essentially PC1 accounts for the errors obtained in prediction ($Dataset_P$), PC2 is related to the errors in cross-validation ($Dataset_{CV}$) and PC3 accounts for the differences between the error metrics. Fig. 4 reports the biplot of PC1 and PC2 feature space, explaining 81.90% of the total variance. Concerning the predictive performance of the learn-based regressors, LR generally led to a higher error in prediction for both ST and MT approaches, while RF is the algorithm giving the best prediction results. Considering the comparison between single-target and multi-target approaches, generally ERC led to the higher enhancement of model performances, except for LR algorithm where the results obtained with DSTARS, ERC and ST are almost the same or inferior (similar to what observed in Fig. 3). In addition, it is possible to highlight that SVM coupled with ERC or DSTARS is the most stable algorithm since it led to lower error values in both cross-validation and prediction.

In order to emphasise a possible superiority of the different combinations of algorithms and MT strategies, Friedman's statistical test and the Nemenyi post hoc test have been applied to the averaged (P dataset) aRRMSE values. Fig. 5 shows the Critical Difference (CD) diagram obtained from the statistical test results. The different models are connected when statistically significant differences are not observed between them ($\alpha = 0.05$ and CD = 7.07). Lower rank values indicate the most accurate (lower aRRMSE) methods, while higher values indicate the less accurate ones.

According to the results reported in Fig. 5, no statistically significant differences were observed when comparing ERC + RF, DSTARS + RF, ERC + SVM, DSTARS + SVM, MTRS + RF, ST + RF, ST + SVM, DSTARS + LR and MTRS + SVM.

Table 1 – Average *aRRMSE* and R^2 performances of ST, ERC, MTRS and DSTARS approaches coupled with linear regression obtained on $Dataset_{CV}$ and $Dataset_P$.

	ST		ERC		MTRS		DSTARS	
	aRRMSE	R^2	aRRMSE	R^2	aRRMSE	R^2	aRRMSE	R^2
$Dataset_{CV}$	0.7224	0.4378	**0.7121**	**0.4541**	0.7238	0.4366	0.7148	0.4506
$Dataset_p$	0.7454	0.4008	0.7366	0.4143	0.7664	0.3680	**0.7304**	**0.4230**

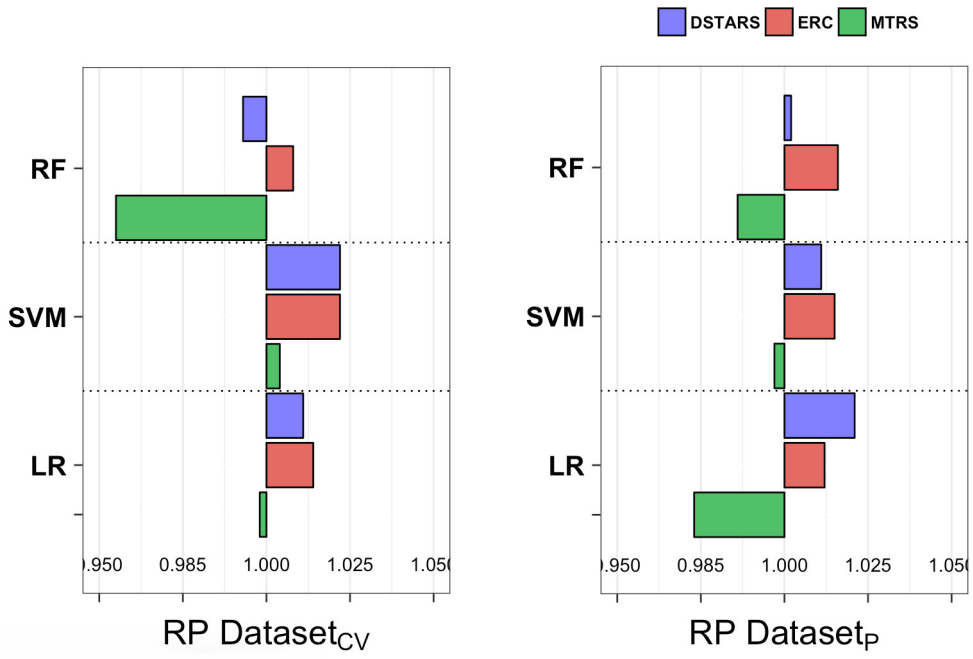

Fig. 3 – Relative Performance (RP) comparison of RF, SVM and LR predictor over all datasets (*Dataset*$_{CV}$ and *Dataset*$_{P}$).

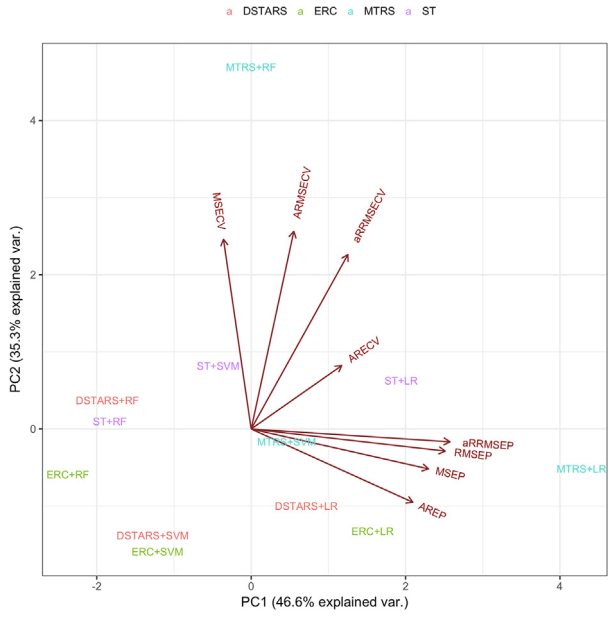

Fig. 4 – PC space obtained from techniques (ST, MTRS, ERC and DSTARS) combined to algorithms (RF, SVM and LR) over all datasets (CV and P) by Mean Squared Error (MSE), Average Root Mean Squared Error (ARMSE), Average Relative Root Mean Squared Error (aRRMSE) and Average Relative Error (ARE).

Nevertheless, comparing the aRRMSE value of each regression algorithm, all the LR solutions obtained inferior results, including the ST + LR (PLS core). MTRS technique did not lead to significant prediction improvements; in some cases, the results obtained with this technique were inferior to those gained using the ST approach, which in turn was a competitive solution when coupled with RF and SVM non-linear algorithms.

3.2. *Algorithms prediction improvements*

MT prediction problems are composed of targets that present different prediction complexity and distinct inter-correlation with each other. Section 2.1 reported the results obtained from Pearson linear correlation between the targets, which showed a considerable variation.

In order to evaluate the error reduction achieved by MT approaches separately for each target, the RP_t values obtained by comparing the *RMSE* performance between MT techniques and ST have been calculated for each different target. The results are shown in Fig. 6, where the RP_t values are reported as a heat map. For each combination of MT approach and lean-based regressor, a reddish colour is related to a decrease of model performance compared the corresponding ST method, while a greenish colour is related to an increase of model performances.

For both datasets, the predictive performances of *Hectolitre Weight* target were enhanced by the use of MT approaches, except for the case of MTRS + RF and MTRS + LR. In these last configurations, the target was predicted with inferior result when compared to ST method with RP_t 0.96 and RP_t 0.99, respectively. Similar behaviour was also observed for *P/L*, which has been improved with respect to ST strategy, and partly for *Stability*, for which improvements were observed only in cross-validation. Conversely, *Protein* and *W* showed lower error reduction, mainly in the case of MTRS, independently from ML algorithm. However, it has to be considered that the Coefficient of Determination of *Protein* target obtained by PLS (ST + LR) was equal to 0.90 (R^2), while for the same target the R^2 value in prediction obtained with DSTARS + RF was equal to 0.91. Therefore, for some targets

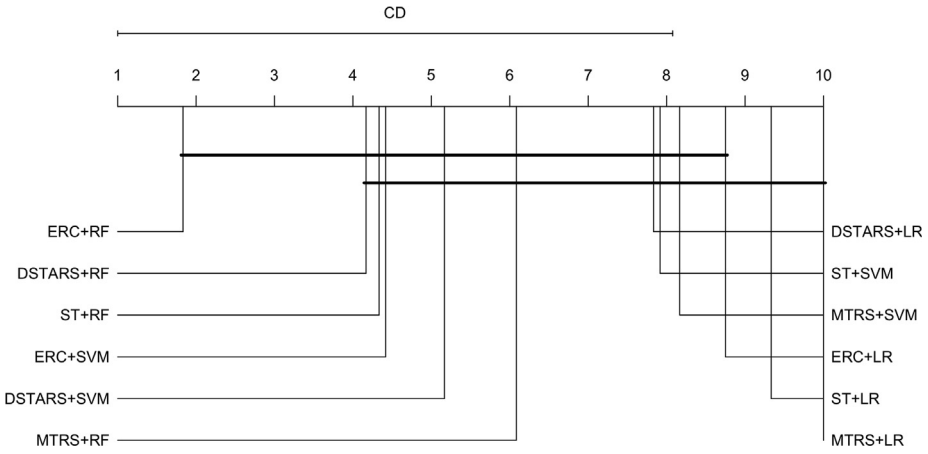

Fig. 5 – Comparison of the averaged aRRMSE values from all technique and algorithm according to the Nemenyi test. Groups of methods that are not significantly different ($\alpha = 0.05$ and CD = 7.07) are connected.

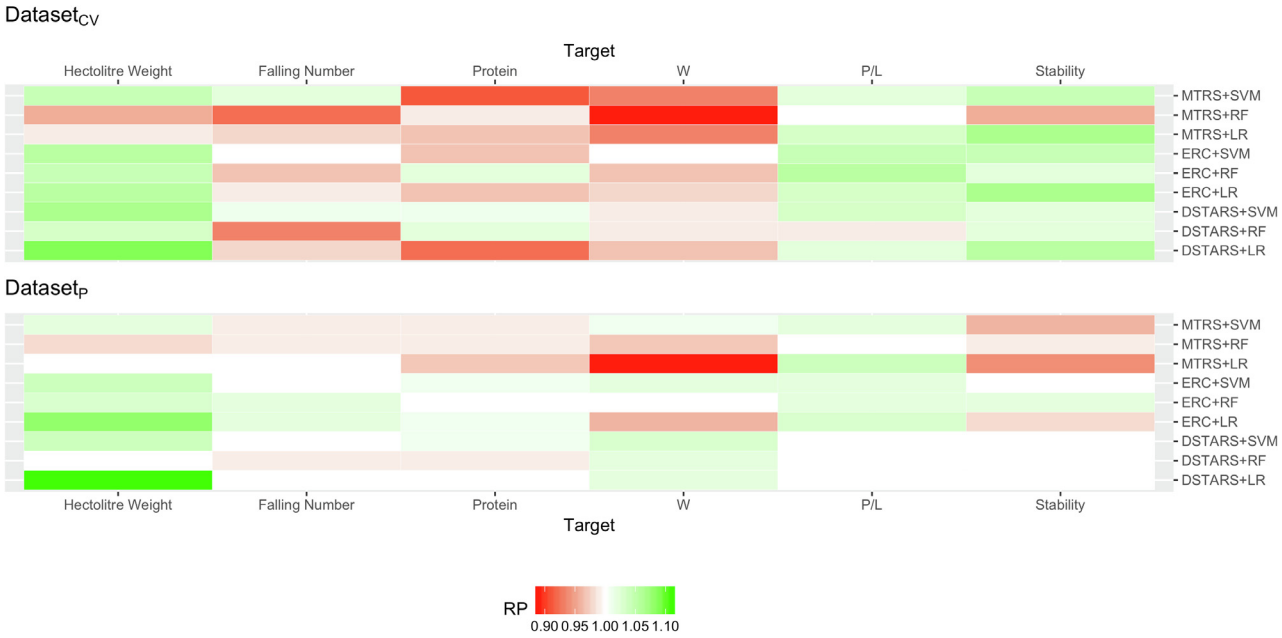

Fig. 6 – Relative Performance improvement of *RMSE* from each target of *Dataset*$_{CV}$ and *Dataset*$_{P}$ between MT techniques and ST.

standard PLS was sufficient to obtain satisfactory results, and the use of MT approaches can lead only to slight improvements.

On the other hand, DSTARS + LR improved the performance of *Hectolitre Weight* from 0.50 (R^2) to 0.56 (R^2). This result meets the importance of exploring different levels of intra-target correlation to perform the target prediction. An important advantage of DSTARS is the possibility to create a MT strategy through the correlation between the targets and to reveal it by the depth of Tracking procedure. For this reason, DSTARS was capable of facing this issue obtaining results superior to ST by dealing with the different inter-correlation between the targets as shown in Fig. 7.

Fig. 7 shows the RMSE values associated with the model layers of each target. In some cases, for example *Protein* and *P/L*, an additional layer is not required to produce good predictors. However, some targets such as *Stability* and *Falling Number* required an additional layer to improve the overall prediction. This figure exposes the RMSE of Modeling and Testing sets split in the kernel of DSTARS strategy as described in Section 2.3.3. The obtained representations meet the heat map representation of Fig. 6, since the targets with smaller improvement from MT approach require a shallow layer structure of DSTARS. This type of investigation was not supported by ERC technique; however, this latter allowed to achieve a better prediction error reduction, as exposed in Section 3.3.

3.3. *Target prediction improvements*

Different prediction improvements were obtained for each target. In Table 2, the RMSE values obtained for all the tested models are reported where, for each target, the best predictive performance is highlighted in bold. For cross-validation, no

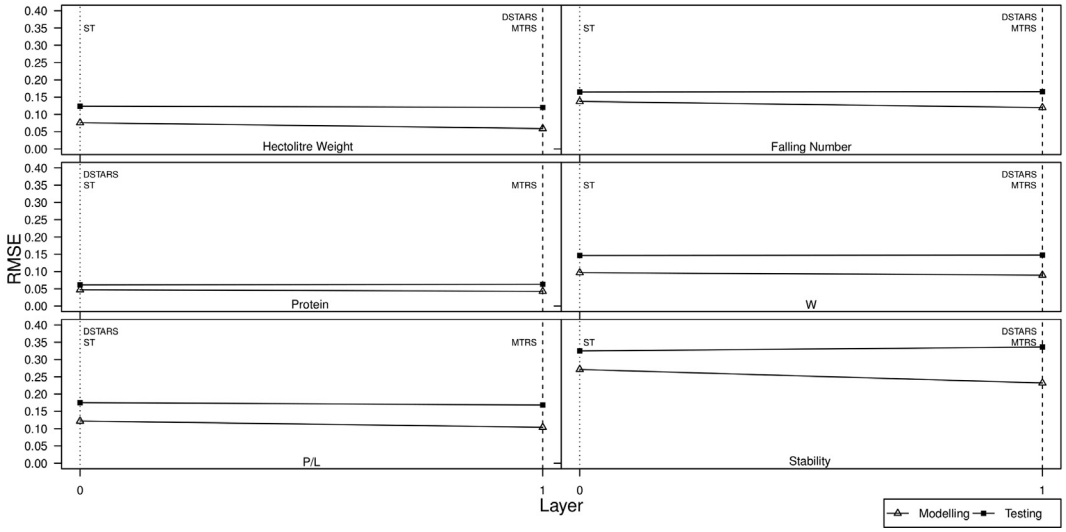

Fig. 7 – DSTARS depth of layers obtained by the use of SVM as base-learner to predict the samples from $Dataset_{CV}$.

Table 2 – RMSE obtained for each target from CV and P datasets by performing all algorithms and techniques.

Dataset	Technique	Algorithm	*RMSE*					
			HlW	FN	Prot	W	P/L	Stab
$Dataset_{CV}$	ST	RF	2.715	56.816	0.406	**66.053**	0.517	5.892
$Dataset_{CV}$	MTRS	RF	2.831	60.835	0.410	73.469	0.515	6.127
$Dataset_{CV}$	ERC	RF	2.615	58.563	0.397	68.285	0.494	5.785
$Dataset_{CV}$	DSTARS	RF	2.634	60.285	0.397	66.810	0.521	5.796
$Dataset_{CV}$	ST	SVM	2.720	57.565	0.420	66.768	0.497	6.072
$Dataset_{CV}$	MTRS	SVM	2.616	**56.449**	0.458	71.179	0.487	5.863
$Dataset_{CV}$	ERC	SVM	2.584	57.401	0.432	66.767	**0.479**	5.831
$Dataset_{CV}$	DSTARS	SVM	**2.563**	56.788	0.417	67.313	0.482	5.930
$Dataset_{CV}$	ST*	LR	2.859	58.744	**0.391**	68.187	0.525	5.824
$Dataset_{CV}$	MTRS	LR	2.875	59.840	0.404	72.922	0.510	5.512
$Dataset_{CV}$	ERC	LR	2.726	59.603	0.403	69.854	0.511	**5.504**
$Dataset_{CV}$	DSTARS	LR	2.656	60.154	0.419	70.096	0.514	5.533
$Dataset_{p}$	ST	RF	2.643	64.064	**0.382**	74.097	0.517	5.514
$Dataset_{p}$	MTRS	RF	2.702	64.593	0.386	76.650	0.515	5.596
$Dataset_{p}$	ERC	RF	**2.558**	**62.935**	0.383	74.080	**0.506**	5.423
$Dataset_{p}$	DSTARS	RF	2.638	64.601	0.387	**72.587**	0.516	**5.498**
$Dataset_{p}$	ST	SVM	2.726	64.045	0.389	75.541	0.532	5.812
$Dataset_{p}$	MTRS	SVM	2.669	64.630	0.394	74.805	0.522	6.051
$Dataset_{p}$	ERC	SVM	2.633	63.978	0.385	73.915	0.519	5.797
$Dataset_{p}$	DSTARS	SVM	2.628	64.045	0.385	73.634	0.532	5.812
$Dataset_{p}$	ST*	LR	2.979	66.759	0.419	74.251	0.532	5.689
$Dataset_{p}$	MTRS	LR	2.994	66.467	0.434	82.968	0.514	6.066
$Dataset_{p}$	ERC	LR	2.769	65.570	0.414	77.101	0.515	5.831
$Dataset_{p}$	DSTARS	LR	2.696	66.759	0.419	72.665	0.532	5.689

* Corresponding to the PLS model results.

one technique significantly outperforms the others. On the other hand, ERC technique turned out to be superior for the majority of targets in prediction set. For $Dataset_{CV}$, SVM algorithm obtained the lowest prediction error for more targets (*HlW*, FN and *P/L*), but in the experiments conducted over $Dataset_{P}$ the RF algorithms obtained the lowest RMSE for all targets.

In order to compare the tested combination of MT approaches and base-learners with usual PLS, the RP values have also been calculated considering ST + LR results as reference. The obtained RP values from $Dataset_{P}$ are reported in Table 3.

All the RP values are superior to usual PLS (RP values greater than 1), except for MTRS built with LR models. This result can be due to a naive MTRS modelling obtained by linear regression, since the linear inter-correlations had already been explored by PLS scores extraction. In other words, all linear dependencies between the targets were explored in the

Table 3 – Relative performance comparison based on usual PLS (ST + LR).

	LR	SVM	RF	Average*
MTRS	0.97	1.02	1.03	1.01
ERC	1.01	1.04	1.07	1.04
DSTARS	1.02	1.03	1.05	1.04
Average**	1.00	1.03	1.05	

* Average calculated from ML algorithms.
** Average calculated from MT techniques.

first step of the proposed approach, and the addition of another linear inference layer does not lead to an improvement. On the other hand, by exploring several different sequential chains and different layers depth, ERC and DSTARS could improve the performances. Also, the combination of non-linear ML algorithms and MT techniques allowed to improve the performance up to 7% (ERC + RF).

Taking into account the results reported in the present study, we suggest the use of Multi-target over PLS scores modelled by a ML algorithm. Our solution was able to treat some drawbacks of PLS by non-linear ML modelling. Indeed, RF could be used to report extra information from RF importance and requires fewer hyper-parameters than SVM. In the case of a demand for high predictive power, ERC is the recommended choice.

4. Conclusion

In the present study, we explored the possibility of taking advantage from multi-target approaches for the simultaneous prediction of different quality-related parameters with NIR spectroscopy. In particular, two different aspects related to the analysis of spectral data have been jointly considered: the benefits of non-linear modelling and the possible advantages of multi-target prediction. Moreover, by means of PLS, the data dimensionality was reduced and a machine learning algorithm was applied to deal with non-linearities in NIR spectra. In addition, the results were further confirmed by a robust validation procedure, allowing to avoid overfitting. Finally, by the use of a Multi-target strategy, it was possible to overcome the actual predictive performance by accounting for the relationships between the targets. Considering all the evaluated quality parameters, this procedure allowed to obtain an increase in the predictive performance up to 7%.

Declaration of Competing Interest

The authors declare they have no significant competing financial, professional, or personal interests that might have influenced the performance or presentation of the work described in this manuscript.

Acknowledgements

The authors would like to thank the financial support of Coordination for the National Council for Scientific and Technological Development (CNPq) of Brazil - Grant of Project 420562/2018-4.

REFERENCES

[1] Aho T, Zenko B, Dzeroski S, Elomaa T. Multi-target regression with rule ensembles. J Mach Learn Res 2012;13:2367–407.
[2] Alishahi A, Farahmand H, Prieto N, Cozzolino D. Identification of transgenic foods using NIR spectroscopy: a review. Spectrochim Acta A: Mol Biomol Spectrosc 2010;75(1):1–7.
[3] Andrés S, Murray I, Navajas EA, Fisher AV, Lambe NR, Bünger L. Prediction of sensory characteristics of lamb meat samples by near infrared reflectance spectroscopy. Meat Sci 2007;76(3):509–16.
[4] Balage JM, Silva SL, Gomide CA, Bonin MN, Figueira AC. Predicting pork quality using Vis/NIR spectroscopy. Meat Sci 2015;108:37–43.
[5] Bao Y, Liu F, Kong W, Sun DW, He Y, Qiu Z. Measurement of soluble solid contents and ph of white vinegars using VIS/NIR spectroscopy and least squares support vector machine. Food Bioprocess Technol 2014;1(7):54–61.
[6] Barbon APAC, Barbon S, Mantovani RG, Fuzyi EM, Peres LM, Bridi AM. Storage time prediction of pork by computational intelligence. Comput Electron Agric 2016;127:368–75.
[7] Ben-Hur A, Weston J. A user's guide to support vector machines. In: Carugo O, Eisenhaber F, editors. Data mining techniques for the life sciences. USA: Humana Press; 2010. p. 223–39.
[8] Blanco M, Villarroya I. NIR spectroscopy: a rapid-response analytical tool. Trends Anal Chem 2002;21(4):240–50.
[9] Borchani H, Varando G, Bielza C, Larrañaga P. A survey on multi-output regression. Wiley Interdiscip Rev: Data Min Knowl Disc 2015;5(5):216–33.
[10] Breiman L. Random forests. Mach Learn 2001;45(1):5–32.
[11] Cocchi M, Corbellini M, Foca G, Lucisano M, Pagani MA, Tassi L, Ulrici A. Classification of bread wheat flours in different quality categories by a wavelet-based feature selection/classification algorithm on NIR spectra. Anal Chim Acta 2005;544(1):100–7.
[12] Cornell JA, Berger RD. Factors that influence the value of the coefficient of determination in simple linear and nonlinear regression models. Phytopathology 1987;77(1):63–70.
[13] Cunha CL, Luna AS, Oliveira RCG, Xavier GM, Paredes MLL, Torres AR. Predicting the properties of biodiesel and its blends using mid-FT-IR spectroscopy and first-order multivariate calibration. Fuel 2017;204:185–94.
[14] De Jong S. Simpls: an alternative approach to partial least squares regression. Chemom Intell Lab Syst 1993;18(3):251–63.
[15] Demšar J. Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 2006;7:1–30.
[16] El-Bendary N, El Hariri E, Hassanien AE, Badr A. Using machine learning techniques for evaluating tomato ripeness. Expert Syst Appl 2015;42(4):1892–905.
[17] Fazayeli A, Kamgar S, Nassiri SM, Fazayeli H, Guardia M. Dielectric spectroscopy as a potential technique for

prediction of kiwifruit quality indices during storage. Inf Process Agric 2019.
[18] Fleureau J, Bensalah K, Rolland D, Lavastre O, Rioux-Leclercq N, Guillé F, Patard JJ, De Crevoisier R, Senhadji L. Characterization of renal tumours based on raman spectra classification. Expert Syst Appl 2011;38(11):14301–6.
[19] Foca G, Ferrari C, Sinelli N, Mariotti M, Lucisano M, Caramanico R, Ulrici A. Minimisation of instrumental noise in the acquisition of FT-NIR spectra of bread wheat using experimental design and signal processing techniques. Anal Bioanal Chem 2011;399(6):1965–73.
[20] Foca G, Salvo D, Cino A, Ferrari C, Fiego DPL, Minelli G, Ulrici A. Classification of pig fat samples from different subcutaneous layers by means of fast and non-destructive analytical techniques. Food Res Int 2013;52(1):185–97.
[21] Foca G, Ulrici A, Corbellini M, Pagani MA, Lucisano M, Franchini GC, Tassi L. Reproducibility of the italian ISQ method for quality classification of bread wheats: an evaluation by expert assessors. J Sci Food Agric 2007;87 (5):839–46.
[22] Galasso HL, Callier MD, Bastianelli D, Blancheton JP, Aliaume C. The potential of near infrared spectroscopy (NIRS) to measure the chemical composition of aquaculture solid waste. Aquaculture 2017;476:134–40.
[23] Geronimo BC, Mastelini SM, Carvalho RH, Barbon S, Barbin DF, Shimokomaki M, Ida EI. Computer vision system and near-infrared spectroscopy for identification and classification of chicken with wooden breast, and physicochemical and technological characterization. Infrared Phys Technol 2019;96:303–10.
[24] Ghasemi JB, Tavakoli H. Application of random forest regression to spectral multivariate calibration. Anal Methods 2013;5(7):1863–71.
[25] Hadavandi E, Shahrabi J, Shamshirband S. A novel Boosted-neural network ensemble for modeling multi-target regression problems. Eng Appl Artif Intell 2015;45:204–19.
[26] Hastie T, Taylor J, Tibshirani R, Walther G. Forward stagewise regression and the monotone lasso. Electron J Stat 2007;1:1–29.
[27] Kim SB, Temiyasathit C, Bensalah K, Tuncel A, Cadeddu J, Kabbani W, Mathker AV, Liu H. An effective classification procedure for diagnosis of prostate cancer in near infrared spectra. Expert Syst Appl 2010;37(5):3863–9.
[28] Kocev D, Džeroski S, White MD, Newell GR, Griffioen P. Using single-and multi-target regression trees and ensembles to model a compound index of vegetation condition. Ecol Model 2009;220(8):1159–68.
[29] Kocev D, Vens C, Struyf J, Džeroski S. Ensembles of multi-objective decision trees. In: Proceedings of European conference on machine learning. Berlin, Heidelberg; 2007. p. 624–31.
[30] Kuang B, Tekin Y, Mouazen AM. Comparison between artificial neural network and partial least squares for on-line visible and near infrared spectroscopy measurement of soil organic carbon, pH and clay content. Soil Tillage Res 2015;146:243–52.
[31] Lee J, Chang K, Jun CH, Cho RK, Chung H, Lee H. Kernel-based calibration methods combined with multivariate feature selection to improve accuracy of near-infrared spectroscopic analysis. Chemom Intell Lab Syst 2015;147:139–46.
[32] Li Vigni M, Durante C, Foca G, Marchetti A, Ulrici A, Cocchi M. Near infrared spectroscopy and multivariate analysis methods for monitoring flour performance in an industrial bread-making process. Anal Chim Acta 2009;642(1–2):69–76.
[33] Liu C, Yang SX, Deng L. A comparative study for least angle regression on NIR spectra analysis to determine internal qualities of navel oranges. Expert Syst Appl 2015;42 (22):8497–503.
[34] Liu G, Lin Z, Yu Y. Multi-output regression on the output manifold. Pattern Recognit 2009;42(11):2737–43.
[35] Mastelini SM, Costa VGT, Santana EJ, Nakano FK, Guido RC, Cerri R, Barbon S. Multi-output tree chaining: an interpretative modelling and lightweight multi-target approach. J Signal Process Syst 2019;91(2):191–215.
[36] Mastelini SM, Santana EJ, Cerri R, Barbon S. DSTARS: a multi-target deep structure for tracking asynchronous regressor stack. In: Proceedings of Brazilian conference on intelligent systems (BRACIS). Uberlandia, Brazil; 2017. p. 216–33.
[37] Melki G, Cano A, Kecman V, Ventura S. Multi-target support vector regression via correlation regressor chains. Inf Sci 2017;415:53–69.
[38] Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, Hamprecht FA. A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinf 2009;10(1):213.
[39] Perez IMN, Badaró AT, Barbon S, Barbon APAC, Pollonio MAR, Barbin DF. Classification of chicken parts using a portable near-infrared (NIR) spectrophotometer and machine learning. Appl Spectrosc 2018;72(12):1774–80.
[40] Porep JU, Kammerer DR, Carle R. On-line application of near infrared (NIR) spectroscopy in food production. Trends Food Sci Technol 2015;46(2):211–30.
[41] Prieto N, Lopez-Campos O, Aalhus JL, Dugan MER, Juarez M, Uttaro B. Use of near infrared spectroscopy for estimating meat chemical composition, quality traits and fatty acid content from cattle fed sunflower or flaxseed. Meat Sci 2014;98(2):279–88.
[42] Ramírez-Morales I, Rivero D, Fernández-Blanco E, Pazos A. Optimization of NIR calibration models for multiple processes in the sugar industry. Chemom Intell Lab Syst 2016;159:45–57.
[43] Santana EJ, Geronimo BC, Mastelini SM, Carvalho RH, Barbin DF, Ida EI, Barbon S. Predicting poultry meat characteristics using an enhanced multi-target regression method. Biosyst Eng 2018;171:193–204.
[44] Santana EJ, Mastelini SM, Barbon S. Deep regressor stacking for air ticket prices prediction. In: Proceedings of XIII Brazilian symposium of information systems: information systems for participatory digital governance (SBSI). Lavras, Brazil; 2017. p. 25–31.
[45] Schmutzler M, Beganovic A, Böhler G, Huck CW. Methods for detection of pork adulteration in veal product based on FT-NIR spectroscopy for laboratory, industrial and on-site analysis. Food Control 2015;57:258–67.
[46] Spyromitros-Xioufis E, Tsoumakas G, Groves W, Vlahavas I. Multi-target regression via input space expansion: treating targets as inputs. Mach Learn 2016;104(1):55–98.
[47] Tange R, Rasmussen MA, Taira E, Bro R. Application of support vector regression for simultaneous modelling of near infrared spectra from multiple process steps. J Near Infrared Spectrosc 2015;23(2):75–84.
[48] Tsoumakas G, Spyromitros-Xioufis E, Vrekou A, Vlahavas I. Multi-target regression via random linear target combinations. In: Proceedings of European conference on machine learning and knowledge discovery in databases. Berlin, Heidelberg; 2014. p. 225–40.
[49] Xiong T, Bao Y, Hu Z. Multiple-output support vector regression with a firefly algorithm for interval-valued stock price index forecasting. Knowl Based Syst 2014;55:87–100.
[50] Xu S, An X, Qiao X, Zhu L, Li L. Multi-output least-squares support vector regression machines. Pattern Recognit Lett 2013;34(9):1078–84.
[51] Zhang LG, Zhang X, Ni LJ, Xue ZB, Gu X, Huang SX. Rapid identification of adulterated cow milk by non-linear pattern

recognition methods based on near infrared spectroscopy. Food Chem 2014;145:342–8.
[52] Zhang W, Liu X, Ding Y, Shi D. Multi-output LS-SVR machine in extended feature space. In: Proceedings of the 2012 IEEE international conference on computational intelligence for measurement systems and applications (CIMSA). Tianjin, China; 2012. p. 130–4.
[53] Zhen X, Yu M, He X, Li S. Multi-target regression via robust low-rank learning. IEEE Trans Pattern Anal Mach Intell 2018;40(2):497–504.